



Entropy-Transport distances between unbalanced metric measure spaces

Nicolò De Ponti¹ · Andrea Mondino² 

Received: 22 September 2020 / Revised: 4 January 2022 / Accepted: 23 July 2022 /
Published online: 10 August 2022
© Crown 2022

Abstract

Inspired by the recent theory of Entropy-Transport problems and by the **D**-distance of Sturm on *normalised* metric measure spaces, we define a new class of complete and separable distances between metric measure spaces of possibly different total mass. We provide several explicit examples of such distances, where a prominent role is played by a geodesic metric based on the Hellinger-Kantorovich distance. Moreover, we discuss some limiting cases of the theory, recovering the “pure transport” **D**-distance and introducing a new class of “pure entropic” distances. We also study in detail the topology induced by such Entropy-Transport metrics, showing some compactness and stability results for metric measure spaces satisfying Ricci curvature lower bounds in a synthetic sense.

Mathematics Subject Classification Primary 54E35; Secondary 49J45 · 53C23 · 60B10

Contents

Introduction	160
1 Preliminaries and notation	165
1.1 Metric and measure setting	165
1.2 Curvature-Dimension condition	169
1.3 Entropy functionals	171
2 Entropy-Transport problem and distances	173
2.1 Regular Entropy-Transport distances	174
3 Sturm-Entropy-Transport distance	175
3.1 Topology	189

✉ Andrea Mondino
Andrea.Mondino@maths.ox.ac.uk
Nicolò De Ponti
ndeponti@sissa.it

¹ Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy

² Mathematical Institute, University of Oxford, Oxford, UK

4 Limiting cases	191
4.1 Pure entropy distances	191
4.2 Sturm's distances	199
4.3 Piccoli-Rossi distance	201
4.4 Bounds between distances	202
5 Comparison with conic Gromov–Wasserstein	202
References	206

Introduction

With motivations from pure Mathematics as well as from applied sciences, over the last decades a growing attention has been paid to the problem of “comparing objects”, which come naturally endowed with a distance/metric and a weight/volume form/measure. From the mathematical point of view, such objects are formalised as metric measure spaces (m.m.s. for short) (X, d, μ) , where the metric structure (X, d) describes the geometry and the mutual distance of points, and the measure μ “weights” the relative importance of different parts of the object.

The flexibility of such a framework allows to unify the treatment of a series of problems stemming from various fields of science and technology, e.g. chemistry [24], data science [33], multi-omics data alignment [13], computer vision [40], language processing [1], graph [46] and shape [42, 49] matching, barycenters & shape analysis [34], generative networks [6, 11], machine learning [47]. The theory of metric measure spaces has been flourishing in pure Mathematics as well, providing a unified setting to investigate concentration of measure phenomena [26, 41], the theory of Ricci limit spaces [8, 19] and, more generally, synthetic notions of Ricci curvature lower bounds [4, 30, 43, 44].

In order to “quantify the similarities and differences between two such objects”, it is thus natural to investigate appropriate notions of distance between metric measure spaces. This idea has its roots in the work of Gromov [23, Chapter 3 $\frac{1}{2}$]), who first recognized the importance of studying the “space of spaces” \mathbf{X} as a metric space in its own right. Formally, \mathbf{X} denotes the set of equivalence classes of metric measure spaces (X, d, μ) , where (X, d) is a complete and separable metric space, and μ is a finite, nonnegative, Borel measure; we are naturally identifying two m.m.s. (X_1, d_1, μ_1) , (X_2, d_2, μ_2) if there exists an isometry $\psi : \text{supp}(\mu_1) \rightarrow \text{supp}(\mu_2)$ such that $\psi_{\#} \mu_1 = \mu_2$. Here by $\text{supp}(\mu)$ we denote the support of the measure μ (see the preliminary section for more details).

In the recent years, the theory has been pushed forward by the works of Sturm [43, 45] and Memoli [32] who realized that ideas from mass transportation can be used to produce new relevant distances between metric measure spaces. Such distances have been successfully applied in different fields, but suffer from a major restriction which is intrinsic of the Wasserstein distances coming from optimal transport: they can be used to compare only spaces *with the same total mass*.

The goal of the present paper is to overcome this limitation by taking advantage of the theory of optimal *Entropy-Transport* problems [29]. In contrast with the classical transport setting, these problems allow the description of phenomena where the conservation of mass may not hold; for this reason they are also known in the literature as

“unbalanced optimal transport problems”. The corresponding theory is fairly recent and is becoming increasingly popular in applications, e.g. gradient flows to train neural networks [9, 37], supervised learning [18], medical imaging [17] and video [27] registration. Indeed, the Entropy-Transport relaxation seems to outperform classical optimal transport in all the problems where the input data is noisy or a normalization procedure is not appropriate. We refer to [38] and references therein for more applications of unbalanced optimal transport.

As we are going to explain in detail below, inspired by the construction of the \mathbf{D} -distance of Sturm [43], we are able to produce a new class of complete and separable distances between metric measure spaces by replacing the Wasserstein distance with an Entropy-Transport distance. Such metric structures on \mathbf{X} also turn out to be geodesic (resp. length) when the underlying Entropy-Transport distance is geodesic (resp. length).

Optimal transport and Sturm distances. Let (X, d) be a metric space and $\mathbf{c} : X \times X \rightarrow [0, +\infty]$ be a lower semi-continuous cost function. The optimal transport problem between two probability measures μ_1, μ_2 consists in the minimization problem:

$$T(\mu_1, \mu_2) := \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int_{X \times X} \mathbf{c}(x_1, x_2) d\gamma(x_1, x_2). \quad (1)$$

Here $\Pi(\mu_1, \mu_2)$ denotes the set of measures γ in the product space $X \times X$ whose marginals satisfy the constraint $\pi_{\#}^i \gamma = \mu_i$, where π^i denotes the projection map $\pi^i(x_1, x_2) = x_i$.

A typical choice for the cost function is $\mathbf{c}(x_1, x_2) = d^p(x_1, x_2)$, $p \geq 1$. In this situation, the transport cost T is the p -power of the celebrated p -Wasserstein distance \mathcal{W}_p , a metric on the set $\mathcal{P}_p(X)$ of probability measures over X with finite p -moment. Starting from the seminal work of Kantorovich, the metric space $(\mathcal{P}_p(X), \mathcal{W}_p)$ has been thoroughly studied: it inherits many geometric properties of the underlying space (X, d) (such as completeness, separability, geodesic property) and induces the weak topology (with p -moments) of probability measures. We refer to the monograph [48] for a detailed overview of the topic.

As observed by Sturm [43], one can lift the metric \mathcal{W}_p to a distance between metric measure spaces by defining:

$$\mathbf{D}_p((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) := \inf \mathcal{W}_p(\psi_{\#}^1 \mu_1, \psi_{\#}^2 \mu_2), \quad (2)$$

where the infimum is taken over all complete and separable metric spaces (\hat{X}, \hat{d}) , and all isometric embeddings $\psi^i : \text{supp}(\mu_i) \rightarrow \hat{X}$. It is proved in [43, Theorem 3.6] that \mathbf{D}_p is a complete, separable and geodesic distance on the set

$$\mathbf{X}_{1,p} := \{(X, d, \mu) \in \mathbf{X} : \mu \in \mathcal{P}_p(X, d)\}.$$

Entropy-Transport problems and Sturm-Entropy-Transport distances. The idea at the core of Entropy-Transport problems is to relax the marginal constraints

typical of the classical Kantorovich formulation (1) by adding some suitable penalizing functionals which keep track of the deviation of the marginals $\gamma_i := \pi_{\sharp}^i \gamma$ from the data μ_i , $i = 1, 2$.

Following the approach of Liero, Mielke and Savaré [29], given a superlinear, convex function $F : [0, +\infty) \rightarrow [0, +\infty]$ such that $F(1) = 0$ (for simplicity here we assume F to be superlinear, see definition (21) for the general case), one considers the entropy functional (also called Csiszár F -divergence [12])

$$D_F : \mathcal{M}(X) \times \mathcal{M}(X) \rightarrow [0, +\infty], \quad D_F(\gamma || \mu) := \begin{cases} \int_X F\left(\frac{d\gamma}{d\mu}\right) d\mu & \text{if } \gamma \ll \mu, \\ +\infty & \text{otherwise.} \end{cases} \quad (3)$$

Here $\mathcal{M}(X)$ denotes the set of finite, nonnegative, Borel measures over X . A classical example is given by the choice $F = U_1(s) := s \ln(s) - s + 1$, that corresponds to the celebrated Kullback-Leibler divergence (note that when γ and μ are probability measures, D_{U_1} coincides with the celebrated Boltzmann-Shannon entropy $\text{Ent}(\rho\mu|\mu) = \int \rho \log \rho d\mu$).

Given $\mu_1, \mu_2 \in \mathcal{M}(X)$, the Entropy-Transport problem induced by the entropy function F and the cost function \mathbf{c} is then defined as

$$\text{ET}(\mu_1, \mu_2) := \inf_{\gamma \in \mathcal{M}(X \times X)} \left\{ \sum_{i=1}^2 D_F(\gamma_i || \mu_i) + \int_{X \times X} \mathbf{c}(x_1, x_2) d\gamma(x_1, x_2) \right\}. \quad (4)$$

We emphasize that the problem (4) makes perfect sense even when $\mu_1(X) \neq \mu_2(X)$.

As in the case of optimal transport problems, it is natural to consider cost functions of the form $\mathbf{c}(x_1, x_2) = \ell(\mathbf{d}(x_1, x_2))$, where \mathbf{d} is a distance on X and $\ell : [0, \infty) \rightarrow [0, \infty]$ is a general function. With a careful choice of the functions F and ℓ (see [14] for a discussion on the metric properties of Entropy-Transport problems), one is able to produce a distance D_{ET} on the space $\mathcal{M}(X)$ by taking a suitable power of the Entropy-Transport cost ET , namely $D_{\text{ET}} = \text{ET}^a$ for a certain $a \in (0, 1]$.

In the paper we introduce the class of *regular* Entropy-Transport distances. The formal definition of this class of distances is given in Definition 2, here we just mention that any regular Entropy-Transport distance D_{ET} is a complete and separable metric on $\mathcal{M}(X)$ of the form $D_{\text{ET}} = \text{ET}^a$, for an Entropy-Transport cost induced by sufficiently regular functions F and ℓ .

For any regular Entropy-Transport distance, the *Sturm-Entropy-Transport* distance D_{ET} between the (equivalence classes of) m.m.s. $(X_1, \mathbf{d}_1, \mu_1)$, $(X_2, \mathbf{d}_2, \mu_2)$ is then defined as

$$D_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) := \inf D_{\text{ET}}(\psi_{\sharp}^1 \mu_1, \psi_{\sharp}^2 \mu_2) \quad (5)$$

where the infimum is taken over all complete and separable metric spaces $(\hat{X}, \hat{\mathbf{d}})$, and all isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$ and $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$.

The main result of the paper (Theorem 2) is that every Sturm-Entropy-Transport distance defines a complete and separable metric structure on \mathbf{X} . Moreover it satisfies the geodesic (resp. length) property if the distance \mathbf{D}_{ET} satisfies the geodesic (resp. length) property on the space of measures.

We also study in detail the notion of convergence induced by such distances, showing that it corresponds to the weak measured-Gromov convergence introduced in [21]. As a consequence, we obtain a compactness result for the class of m.m.s. (X, d, μ) satisfying the $\text{CD}(K, N)$ condition, having bounded diameter and satisfying $0 < v \leq \mu(X) \leq V$. We refer to Theorem 4 for the precise statement and to the preliminaries for the definition of the curvature-dimension condition $\text{CD}(K, N)$.

At a technical level, the proofs of our results are inspired by the corresponding ones given by Sturm in [43], but they require new ideas in order to deal with general cost functions and with the entropic part of the problem. Two key results of independent interest are contained in Proposition 2 and Lemma 6, where we show that the infimum in the right hand side of (5) is actually a minimum, and we give an explicit formulation of the Sturm-Entropy-Transport distance, namely

$$\mathbf{D}_{\text{ET}}^{1/a}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) = \sum_{i=1}^2 D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{d}(x, y)) d\gamma, \quad (6)$$

for some optimal measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ and optimal pseudo-metric coupling \hat{d} between d_1 and d_2 (see the preliminaries for the definition of pseudo-metric coupling). Also the proof of one of the main results, Theorem 2, despite being inspired by [43], departs from it and needs some new ideas:

- in order to show that \mathbf{D}_p defines a *non-degenerate* distance function (i.e.

$$\mathbf{D}_p((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) = 0$$

implies that (X_1, d_1, μ_1) and (X_2, d_2, μ_2) are isomorphic as metric measure spaces), Sturm [43] establishes a comparison result with Gromov's \square_1 distance, of independent interest; this permits to inherit the non-degeneracy of \mathbf{D}_p by the one of \square_1 .

Instead, we argue directly: thanks to the aforementioned Proposition 2 and Lemma 6, we can exploit the existence of an optimal coupling *both at the level of space and measure* and infer the non-degeneracy of \mathbf{D}_{ET} directly;

- in order to show that the \mathbf{D}_p distance is *length*, Sturm [43] argues by approximation via finite metric spaces, taking advantage of the “pure transport” behaviour of \mathbf{D}_p .

Due to the entropy contribution in the \mathbf{D}_{ET} distance, we argue differently: the main point is to embed everything in a complete, separable and *geodesic* ambient space, obtained by a slight modification of Kuratowski embedding.

The class of regular Entropy-Transport distances includes some of the main examples of Entropy-Transport distances known in the literature, including:

- The Hellinger-Kantorovich geodesic distance [10, 25, 28, 29] induced by the choices

$$a = 1/2, \quad F(s) = U_1(s), \quad \ell(d) = \begin{cases} -\log(\cos^2(d)) & \text{if } d < \frac{\pi}{2}, \\ +\infty & \text{otherwise.} \end{cases}$$

- The so-called *Gaussian* Hellinger-Kantorovich distance [29] that corresponds to the choices

$$a = 1/2, \quad F(s) = U_1(s), \quad \ell(d) = d^2.$$

- The *quadratic power-like* distances studied in [14] corresponding to

$$a = 1/2, \quad F(s) = U_p(s) := \frac{s^p - p(s-1) - 1}{p(p-1)}, \quad \ell(d) = d^2, \quad 1 < p \leq 3.$$

Moreover, our analysis is not restricted to regular Entropy-Transport distances. By a limit procedure we also discuss some singular cases covering:

- The “pure entropy” setting that corresponds to the choice

$$\mathbf{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases}$$

In this situation we construct a family of distances between metric measure spaces inducing a notion of *strong* convergence (see Theorems 5 and 6 for the details).

- The “pure transport” setting, corresponding to

$$a = 1/p, \quad F(s) = \begin{cases} 0 & \text{if } s = 1, \\ +\infty & \text{otherwise,} \end{cases} \quad \ell(d) = d^p,$$

where we recover the \mathbf{D}_p -distances introduced by Sturm.

- The *Piccoli-Rossi* distance BL [35, 36] (also known as *bounded-Lipschitz distance*), induced by the choices

$$a = 1, \quad F(s) = |s - 1|, \quad \ell(d) = d.$$

By an analogous procedure to the one described in (5), in Theorem 8 we show that the distance BL can be lifted to a complete distance \mathbf{BL} on the set \mathbf{X} .

Note on the preparation. Some of the results of the paper (often under additional assumptions) have been presented at different seminars and included in the Phd thesis of the first named author [15, Chapter 5], where the construction of the Sturm-Entropy-Transport distances induced by the Hellinger-Kantorovich and the quadratic power-like distances is developed.

Only during the final stage of preparation of the present manuscript (September 2020), we became aware of the independent work [39], which defines a class of distances (denoted by CGW, for “conic Gromov-Wasserstein”) between unbalanced metric measure spaces starting from the construction of the Gromov-Wasserstein distance introduced in [32] and the *conical* formulation of the Entropy-Transport problems (see [10, 14, 29] and Remark 1 for a discussion on the “cone geometry” of Entropy-Transport problems). The paper [39] also provides some interesting numerical discussions on the topic, while it is not present a study on the analytic and geometric properties of this class of distances (such as completeness, separability, the length and geodesic property, compactness). An expert reader will notice that our \mathbf{D}_{ET} distance is an *unbalanced* counterpart of Sturm’s \mathbf{D}_p distance [43] for probability metric measure spaces, while Séjourné-Vialard-Peyré CGW distance [39] is an *unbalanced* counterpart of Memoli’s [32] Gromov-Wasserstein distance. A major difference between the two approaches is that while our \mathbf{D}_{ET} distance is complete (see Theorem 2), the Gromov-Wasserstein distance of [32] is not complete, and the same is expected for the CGW distance of [39]. The relation between \mathbf{D}_{ET} and CGW is analysed in Sect. 5, where we prove an upper bound of the latter in terms of the former.

1 Preliminaries and notation

1.1 Metric and measure setting

A function $d : X \times X \rightarrow [0, \infty]$ is a *pseudo-metric* on the set X if d is symmetric, satisfies the triangle inequality and $d(x, x) = 0$ for every $x \in X$. We say that d is a metric possibly attaining the value $+\infty$ if it is a pseudo-metric such that $d(x, y) = 0$ implies $x = y$. When d is also finite-valued, we simply say that d is a metric. A *pseudo-metric space* (resp. *metric space*) will be a couple (X, d) , where d is a pseudo-metric (resp. metric) on the set X .

On a pseudo-metric space we will always consider the topology induced by the open balls $B_r(x) := \{y \in X : d(x, y) < r\}$. A *Polish space* is a separable completely metrizable topological space. We will denote by $\text{diam}(X)$ the diameter of a metric space X .

An *isometry* between two metric spaces (X_1, d_1) , (X_2, d_2) is a map $\psi : X_1 \rightarrow X_2$ such that for every $x, y \in X_1$ we have

$$d_1(x, y) = d_2(\psi(x), \psi(y)). \quad (7)$$

Let $\{(X_\alpha, d_\alpha) | \alpha \in A\}$ be an indexed family of metric spaces, we define its *disjoint union* as

$$\bigsqcup_{\alpha} X_{\alpha} := \bigcup \{X_{\alpha} \times \{\alpha\} | \alpha \in A\},$$

endowed with a pseudo-metric \hat{d} , called *pseudo-metric coupling* between $\{d_\alpha\}$, such that $\hat{d}((x, \alpha), (y, \alpha)) = d_\alpha(x, y)$ for every $x, y \in X_\alpha$. The inclusion map

$$\iota_\alpha : X_\alpha \rightarrow \bigsqcup_\alpha X_\alpha, \quad \iota_\alpha(x) := (x, \alpha),$$

is thus an isometry with image $X_\alpha \times \{\alpha\}$. We will often identify, with a slight abuse of notation, the space X_α with $X_\alpha \times \{\alpha\}$.

Lemma 1 *Let (X_1, d_1) , (X_2, d_2) be two complete and separable metric spaces. Let \hat{d} be a finite valued pseudo-metric coupling between d_1 and d_2 . Then the space*

$$\tilde{X} := (X_1 \sqcup X_2) / \sim \quad \text{where} \quad x_1 \sim x_2 \iff \hat{d}(x_1, x_2) = 0 \quad (8)$$

endowed with the distance

$$\tilde{d}([x_1], [x_2]) := \hat{d}(x_1, x_2)$$

is a complete and separable metric space. Here $[x] \in \tilde{X}$ denotes the equivalence class of the point $x \in X_1 \sqcup X_2$.

Proof We firstly notice that \tilde{d} is well defined on \tilde{X} . Indeed, if $x_1 \sim \tilde{x}_1$ and $x_2 \sim \tilde{x}_2$ we have

$$\begin{aligned} \tilde{d}(x_1, x_2) &= \hat{d}(x_1, x_2) \leq \hat{d}(x_1, \tilde{x}_1) + \hat{d}(\tilde{x}_1, \tilde{x}_2) + \hat{d}(\tilde{x}_2, x_2) = \tilde{d}(\tilde{x}_1, \tilde{x}_2) \\ \tilde{d}(\tilde{x}_1, \tilde{x}_2) &= \hat{d}(\tilde{x}_1, \tilde{x}_2) \leq \hat{d}(\tilde{x}_1, x_1) + \hat{d}(x_1, x_2) + \hat{d}(x_2, \tilde{x}_2) = \tilde{d}(x_1, x_2) \end{aligned}$$

which implies $\tilde{d}(x_1, x_2) = \tilde{d}(\tilde{x}_1, \tilde{x}_2)$.

It is clear that \tilde{d} is a metric on $(X_1 \sqcup X_2) / \sim$.

The separability is a consequence of the fact that $(X_1 \sqcup X_2, \hat{d})$ is separable, being the union of two separable space (recall that $\hat{d} = d_i$ on X_i , $i = 1, 2$).

To prove the completeness, let us consider a Cauchy sequence $\{y_j\} \in \tilde{X}$. It is sufficient to show that a subsequence is converging with respect to \tilde{d} . Let $p : X_1 \sqcup X_2 \rightarrow \tilde{X}$ be the quotient map and, recalling that $X_1 \sqcup X_2 = X_1 \times \{0\} \cup X_2 \times \{1\}$, we can suppose without loss of generality that there exists a subsequence $\{p^{-1}(y_{j_k})\} \in X_1 \times \{0\}$ (the case $\{p^{-1}(y_{j_k})\} \in X_2 \times \{1\}$ being analogous). Up to identifying $(X_1 \times \{0\}, \hat{d})$ with (X_1, d_1) , we can infer that $\{p^{-1}(y_{j_k})\}$ is a Cauchy sequence in the complete space (X_1, d_1) and thus it converges. It is immediate to check that $\{y_{j_k}\}$ is converging in \tilde{X} with respect to \tilde{d} and the proof is complete. \square

Starting from a metric space (X, d) , we define the *cone over X* as the space

$$\begin{aligned} \mathfrak{C}(X) &:= (X \times [0, +\infty)) / \sim \quad \text{where} \quad (x_1, r_1) \sim (x_2, r_2) \\ &\iff r_1 = r_2 = 0 \text{ or } r_1 = r_2, x_1 = x_2. \end{aligned}$$

If (X, d) is a pseudo-metric space, we denote by $\mathcal{M}(X)$ the space of finite, non-negative measures on the Borel σ -algebra $\mathcal{B}(X)$, and by $\mathcal{P}(X) \subset \mathcal{M}(X)$ the space

of probability measures. We endow $\mathcal{M}(X)$ with the weak topology, inducing the following notion of convergence:

$$\mu_n \rightarrow \mu \iff \int_X f d\mu_n \rightarrow \int_X f d\mu \text{ for any } f \in C_b(X), \quad (9)$$

where $C_b(X)$ denotes the set of real, continuous and bounded functions defined on X .

A subset $\mathcal{K} \subset \mathcal{M}(X)$ is *bounded* if $\sup_{\mu \in \mathcal{K}} \mu(X) < \infty$ and it is *equally tight* if

$$\forall \epsilon > 0 \exists K_\epsilon \subset X \text{ compact} : \forall \mu \in \mathcal{K}, \mu(X \setminus K_\epsilon) \leq \epsilon. \quad (10)$$

Compactness properties with respect to the weak topology on $\mathcal{M}(X)$ are guaranteed by the following version of Prokhorov's Theorem:

Theorem 1 *Let X be a Polish space. A subset $\mathcal{K} \subset \mathcal{M}(X)$ is bounded and equally tight if and only if it is relatively compact with respect to the weak topology.*

We recall that the set of measures of the form

$$\mu = M \sum_{n=1}^N \delta_{x_n}, \quad (11)$$

where $M \in \mathbb{R}_+$, $N \in \mathbb{N}$ and $x_n \in X$, is dense in $\mathcal{M}(X)$. Moreover, if X is separable, the measures of the form (11), with $M \in \mathbb{Q}_+$ and x_n in a countable dense subset of X , form a countable dense subset of $\mathcal{M}(X)$, proving that also the latter is a separable space.

A *metric measure space* will be a triple (X, d, μ) where (X, d) is a complete, separable metric space and $\mu \in \mathcal{M}(X)$. If there exists a point $x_0 \in X$ such that

$$\int_X d^p(x_0, x) d\mu(x) < \infty, \quad (12)$$

we will say that the measure $\mu \in \mathcal{P}(X)$ has finite p -moment. We denote by $\mathcal{P}_p(X)$ the space of measures $\nu \in \mathcal{P}(X)$ with finite p -moment.

The support of the measure μ is the smallest closed set $X_0 := \text{supp}(\mu)$ such that $\mu(X \setminus X_0) = 0$. We notice that the set $\text{supp}(\mu)$ has a natural structure of metric measure space with the induced distance, σ -algebra and measure (which will be denoted in the same way).

We say that φ is a *curve* connecting $x, y \in X$, if $\varphi : [a, b] \rightarrow X$ is a continuous map such that $\varphi(a) = x$ and $\varphi(b) = y$. The length of a curve is defined as

$$\text{Length}(\varphi) := \sup \sum_{i=1}^n d(\varphi(t_{i-1}), \varphi(t_i)), \quad (13)$$

where the supremum is taken over all the partitions $a = t_0 < t_1 < \dots < t_n = b$.

We will always assume that a curve of finite length is parametrized by constant speed, i.e.

$$\text{Length}(\varphi|_{[s,t]}) = \frac{t-s}{b-a} \text{Length}(\varphi). \quad (14)$$

A metric space (X, d) is called *length space* if for all $x, y \in X$

$$d(x, y) = \inf \{ \text{Length}(\varphi) : \varphi \text{ curve connecting } x \text{ and } y \}. \quad (15)$$

A *geodesic* is a curve $\varphi : [a, b] \rightarrow X$ such that

$$d(\varphi(a), \varphi(t)) = \frac{t-a}{b-a} d(\varphi(a), \varphi(b)), \quad \text{for all } t \in [a, b].$$

Notice in particular that if φ is a geodesic then

$$\text{Length}(\varphi) = d(\varphi(a), \varphi(b)).$$

A metric space (X, d) is geodesic if any pair of points $x, y \in X$ is connected by a geodesic.

For a metric space (X, d) , the Kantorovich-Wasserstein distance \mathcal{W}_p of order p , $p \geq 1$, is defined as follows: for $\mu_0, \mu_1 \in \mathcal{M}(X)$ we set

$$\mathcal{W}_p^p(\mu_0, \mu_1) := \inf_{\gamma} \int_{X \times X} d^p(x, y) d\gamma, \quad (16)$$

where the infimum is taken over all $\gamma \in \mathcal{M}(X \times X)$ with μ_0 and μ_1 as the first and the second marginal, i.e. $(\pi^i)_\# \gamma = \mu_i$ where $\pi^i : X \times X \rightarrow X$ denotes the projection map $\pi^i(x_1, x_2) = x_i$, $i = 1, 2$. A measure $\gamma \in \mathcal{M}(X \times X)$ achieving the minimum in (16) with given marginals is said a \mathcal{W}_p -optimal coupling for (μ_0, μ_1) . It is clear that $\mathcal{W}_p(\mu_1, \mu_2) = +\infty$ when $\mu_1(X) \neq \mu_2(X)$.

If (X, d) is complete and separable, $(\mathcal{P}_p(X), \mathcal{W}_p)$ is a complete and separable metric space. It is geodesic when (X, d) is geodesic. Moreover, for any sequence $\mu_n \in \mathcal{P}_p(X)$ we have

$$\lim_{n \rightarrow \infty} \mathcal{W}_p(\mu_n, \mu) = 0 \iff \begin{cases} \mu_n \text{ weakly converges to } \mu, \\ \mu_n \text{ has uniformly } p\text{-integrable moments,} \end{cases} \quad (17)$$

where the latter means that for some (thus any) x_0

$$\lim_{R \rightarrow \infty} \limsup_n \int_{X \setminus B_R(x_0)} d^p(x_0, x) d\mu_n(x) = 0. \quad (18)$$

For a proof of these last facts, see [48, Theorem 6.18].

1.2 Curvature-Dimension condition

It is out of the scopes of this brief section to give a full account of the curvature-dimension condition and its properties; we will limit to schematically recalling the basic definitions involved. The interested reader is referred to the original papers [3–5, 7, 16, 20, 21, 30, 43, 44], the survey [2] and the monograph [48].

- For any $K \in \mathbb{R}$, $N \in (1, \infty)$, $\theta > 0$ and $t \in [0, 1]$, define the distortion coefficients by

$$\tau_{K,N}^{(t)}(\theta) := t^{\frac{1}{N}} \sigma_{K,N-1}^{(t)}(\theta)^{\frac{N-1}{N}},$$

where

$$\sigma_{K,N}^{(t)}(\theta) := \begin{cases} \infty & \text{if } K\theta^2 \geq N\pi^2 \\ \frac{\sin(t\theta\sqrt{K/N})}{\sin(\theta\sqrt{K/N})} & \text{if } 0 < K\theta^2 < N\pi^2 \\ t & \text{if } K\theta^2 = 0 \\ \frac{\sinh(t\theta\sqrt{K/N})}{\sinh(\theta\sqrt{K/N})} & \text{if } K\theta^2 < 0 \end{cases}.$$

- For every $N \in (1, \infty)$, define the N -Rényi entropy functional relative to μ , $\mathcal{U}_N(\cdot|\mu) : \mathcal{P}(X) \rightarrow [-\infty, 0]$ as

$$\mathcal{U}_N(v|\mu) := - \int_X \rho^{1-\frac{1}{N}} d\mu, \quad \text{where } v = \rho\mu + v^s \text{ and } v^s \perp \mu.$$

- Define also the Boltzmann-Shannon entropy functional relative to μ , $\text{Ent}(\cdot|\mu) : \mathcal{P}(X) \rightarrow (-\infty, +\infty]$ as

$$\text{Ent}(v|\mu) := \int_X \rho \log(\rho) d\mu, \quad \text{if } v = \rho\mu \ll \mu \text{ and } \rho \log \rho \in L^1(X, \mu),$$

and $+\infty$ otherwise.

- $\text{CD}(K, \infty)$ condition: given $K \in \mathbb{R}$, we say that (X, d, μ) verifies the $\text{CD}(K, \infty)$ condition if for any pair of probability measures $\nu_0, \nu_1 \in \mathcal{P}_2(X)$ with

$$\text{Ent}(\nu_0|\mu), \text{Ent}(\nu_1|\mu) < +\infty,$$

there exists a \mathcal{W}_2 -geodesic $(\nu_t)_{t \in [0,1]}$ from ν_0 to ν_1 such that

$$\text{Ent}(\nu_t|\mu) \leq (1-t) \text{Ent}(\nu_0|\mu) + t \text{Ent}(\nu_1|\mu) - \frac{K}{2} t(1-t) \mathcal{W}_2^2(\mu_0, \mu_1),$$

for any $t \in [0, 1]$.

- $\text{CD}(K, N)$ condition: given $K \in \mathbb{R}$, $N \in (1, \infty)$ we say that (X, d, μ) verifies the $\text{CD}(K, N)$ condition if for any pair of probability measures $\nu_0, \nu_1 \in \mathcal{P}_2(X)$ with

bounded support and with $v_0, v_1 \ll \mu$, there exists a \mathcal{W}_2 -geodesic $(v_t)_{t \in [0,1]}$ from v_0 to v_1 with $v_t \ll \mu$, and a \mathcal{W}_2 -optimal coupling $\gamma \in \mathcal{P}(X \times X)$ such that

$$\mathcal{U}_{N'}(v_t | \mu) \leq - \int \left[\tau_{K,N'}^{(1-t)}(d(x, y)) \rho_0^{-\frac{1}{N'}} + \tau_{K,N'}^{(t)}(d(x, y)) \rho_1^{-\frac{1}{N'}} \right] d\gamma(x, y),$$

for any $N' \geq N, t \in [0, 1]$.

- *Consistency property:* A smooth Riemannian manifold (resp. weighted Riemannian manifold) M satisfies the $\text{CD}(K, N)$ condition for some $K \in \mathbb{R}, N \in (1, \infty)$ if and only if $\dim(M) \leq N$ and the Ricci curvature is bounded below by K (resp. if and only if the N -Bakry-Émery-Ricci tensor is bounded below by K).
- Define the *slope* of a real valued function $u : X \rightarrow \mathbb{R}$ at the point $x \in X$ as

$$|\nabla u|(x) := \begin{cases} \limsup_{y \rightarrow x} \frac{|u(x) - u(y)|}{d(x, y)} & \text{if } x \text{ is not isolated} \\ 0 & \text{otherwise.} \end{cases}$$

We denote with $\text{LIP}(X)$ the space of Lipschitz functions on (X, d) .

- Let $f \in L^2(X, \mu)$. The *Cheeger energy* of f is defined as

$$\text{Ch}(f) := \inf \left\{ \liminf_{n \rightarrow \infty} \frac{1}{2} \int |\nabla f_n|^2 d\mu \mid f_n \in \text{LIP}(X) \cap L^2(X, \mu), \|f_n - f\|_{L^2} \rightarrow 0 \right\}.$$

One can check that the Cheeger energy $\text{Ch} : L^2(X, \mu) \rightarrow [0, \infty]$ is convex and lower semi-continuous. Thus it admits an L^2 -gradient flow, called *heat flow*.

- The metric measure space (X, d, μ) is said *infinitesimally Hilbertian* if Ch is a quadratic form, i.e. it satisfies the parallelogram identity.

One can check that (X, d, μ) is infinitesimally Hilbertian if and only if the heat flow for every positive time is a linear map from $L^2(X, \mu)$ to $L^2(X, \mu)$.

If (X, d, μ) is the metric measure space associated to a smooth Finsler manifold, one can check that (X, d, μ) is infinitesimally Hilbertian if and only if the manifold is actually Riemannian.

- Given $K \in \mathbb{R}$ and $N \in (1, \infty]$, we say that (X, d, μ) verifies the $\text{RCD}(K, N)$ condition if it satisfies the $\text{CD}(K, N)$ condition and it is infinitesimally Hilbertian.
- *Pointed measured Gromov-Hausdorff convergence:* Let (X_n, d_n, μ_n) , $n \in \mathbb{N} \cup \{\infty\}$, be a sequence of metric measure spaces and let $\bar{x}_n \in X_n$ for every $n \in \mathbb{N} \cup \{\infty\}$ be a sequence of reference points. We say that $(X_n, d_n, \mu_n, \bar{x}_n) \rightarrow (X_\infty, d_\infty, \mu_\infty, \bar{x}_\infty)$ in the pointed measured Gromov Hausdorff (pmGH) sense, provided for any $\varepsilon, R > 0$ there exists $N(\varepsilon, R) \in \mathbb{N}$ such that for all $n \geq N(\varepsilon, R)$ there exists a Borel map $f_n^{R, \varepsilon} : B_R(\bar{x}_n) \rightarrow X_\infty$ such that

- $f_n^{R, \varepsilon}(\bar{x}_n) = \bar{x}_\infty$,
- $\sup_{x, y \in B_R(\bar{x}_n)} |d_n(x, y) - d_\infty(f_n^{R, \varepsilon}(x), f_n^{R, \varepsilon}(y))| \leq \varepsilon$,
- the ε -neighbourhood of $f_n^{R, \varepsilon}(B_R(\bar{x}_n))$ contains $B_{R-\varepsilon}(\bar{x}_\infty)$,

- $(f_n^{R,\varepsilon})_{\#}(\mu_n \llcorner B_R(\bar{x}_n))$ weakly converges to $\mu_{\infty} \llcorner B_R(x_{\infty})$ as $n \rightarrow \infty$, for a.e. $R > 0$.

If in addition there exists $\bar{R} > 0$ such that $\text{diam}(X_n) \leq \bar{R}$ for every $n \in \mathbb{N} \cup \{\infty\}$, then we say that $(X_n, \mathbf{d}_n, \mu_n) \rightarrow (X_{\infty}, \mathbf{d}_{\infty}, \mu_{\infty})$ in the *measured Gromov Hausdorff (mGH for short)* sense. In this case it is enough to consider only $R = \bar{R}$ in the above requirements.

- **Stability:** Let $K \in \mathbb{R}$ and $N \in (1, \infty]$ be given. Assume that $(X_n, \mathbf{d}_n, \mu_n)$ satisfies $\text{CD}(K, N)$ (resp. $\text{RCD}(K, N)$), for every $n \in \mathbb{N}$, and that $(X_n, \mathbf{d}_n, \mu_n, \bar{x}_n) \rightarrow (X_{\infty}, \mathbf{d}_{\infty}, \mu_{\infty}, \bar{x}_{\infty})$ in the pmGH sense. Then $(X_{\infty}, \mathbf{d}_{\infty}, \mu_{\infty})$ satisfies $\text{CD}(K, N)$ (resp. $\text{RCD}(K, N)$) as well.

1.3 Entropy functionals

In this section we assume that X is a Polish space.

A function $F : [0, +\infty) \rightarrow [0, +\infty]$ belongs to the class $\Gamma_0(\mathbb{R}_+)$ of the *admissible entropy functions* if F is convex, lower semicontinuous and $F(1) = 0$. We define the recession constant as

$$F'_{\infty} = \lim_{s \rightarrow \infty} \frac{F(s)}{s},$$

and we say that F is superlinear if $F'_{\infty} = +\infty$.

We also define the perspective function induced by $F \in \Gamma_0(\mathbb{R}_+)$ as the function $\hat{F} : [0, +\infty) \times [0, +\infty) \rightarrow [0, +\infty]$, given by

$$\hat{F}(r, t) := \begin{cases} F\left(\frac{r}{t}\right)t & \text{if } t > 0, \\ F'_{\infty}r & \text{if } t = 0. \end{cases} \quad (19)$$

The function

$$R : [0, +\infty) \rightarrow [0, +\infty], \quad R(t) := \hat{F}(1, t) \quad (20)$$

is called *reverse entropy*.

Let $F \in \Gamma_0(\mathbb{R}_+)$ be an admissible entropy function. The *F-divergence* (also called *Csiszár's divergence* or *relative entropy*) is the functional $D_F : \mathcal{M}(X) \times \mathcal{M}(X) \rightarrow [0, +\infty]$ defined by

$$D_F(\gamma || \mu) := \int_X F(\sigma) d\mu + F'_{\infty} \gamma^{\perp}(X), \quad \gamma = \sigma \mu + \gamma^{\perp}, \quad (21)$$

where $\gamma = \sigma \mu + \gamma^{\perp}$ is the Lebesgue's decomposition of the measure γ with respect to μ . When F is superlinear $D_F(\gamma || \mu) = +\infty$ if γ has a singular part with respect to μ . Moreover, it is clear that $D_F(\mu || \mu) = 0$.

We now collect some useful properties of the relative entropies. For the proof see [29, Sect. 2.4].

Lemma 2 *The functional D_F is jointly convex and lower semicontinuous in $\mathcal{M}(X) \times \mathcal{M}(X)$. More generally, if $F \in \Gamma_0(\mathbb{R}_+)$ is the pointwise limit of an increasing sequence $(F_n) \subset \Gamma_0(\mathbb{R}_+)$ and $\gamma, \mu \in \mathcal{M}(X)$ are the weak limit of a sequence $(\gamma_n, \mu_n) \subset \mathcal{M}(X) \times \mathcal{M}(X)$ then we have*

$$\liminf D_{F_n}(\gamma_n || \mu_n) \geq D_F(\gamma || \mu).$$

Lemma 3 *If $\mathcal{K} \subset \mathcal{M}(X)$ is bounded and $F'_\infty > 0$ then the set*

$$\mathbf{K}_C := \{\gamma \in \mathcal{M}(X) : D_F(\gamma || \mu) \leq C, \text{ for some } \mu \in \mathcal{K}\} \quad (22)$$

is bounded for every $C \geq 0$. Moreover, if \mathcal{K} is also equally tight and F is superlinear, then \mathbf{K}_C is equally tight for every $C \geq 0$.

The last lemma of this section shows an invariance result for the F -divergences.

Lemma 4 *Let $F \in \Gamma_0(\mathbb{R}_+)$ be an admissible entropy function, X, Y be two Polish spaces and $f : X \rightarrow Y$ be a Borel injective map. Then, for any $\gamma, \mu \in \mathcal{M}(X)$ it holds*

$$D_F(\gamma || \mu) = D_F(f_{\#}\gamma || f_{\#}\mu). \quad (23)$$

Proof Let us consider the Lebesgue's decompositions

$$\gamma = \sigma\mu + \gamma^\perp \quad \text{and} \quad f_{\#}\gamma = \tilde{\sigma}f_{\#}\mu + \tilde{\gamma}^\perp.$$

Since $f_{\#}\gamma$ and $f_{\#}\mu$ have support contained in $f(X)$, we can suppose without loss of generality that f is bijective.

For any Borel set $A \subset X$ we have

$$\begin{aligned} \int_A \sigma \, d\mu + \gamma^\perp(A) &= \gamma(A) = \gamma(f^{-1}(f(A))) = f_{\#}\gamma(f(A)) \\ &= \int_{f(A)} \tilde{\sigma} \, df_{\#}\mu + \tilde{\gamma}^\perp(f(A)) = \int_A \tilde{\sigma} \circ f \, d\mu + \tilde{\gamma}^\perp(f(A)). \end{aligned} \quad (24)$$

By the uniqueness of the Lebesgue's decomposition (see [29, Lemma 2.3]) it follows that $\sigma = \tilde{\sigma} \circ f$ up to $(\mu + \gamma)$ -negligible sets and $\gamma^\perp(X) = \tilde{\gamma}^\perp(f(X)) = \tilde{\gamma}^\perp(Y)$. In particular

$$\begin{aligned} D_F(f_{\#}\gamma || f_{\#}\mu) &:= \int_Y F(\tilde{\sigma}) \, df_{\#}\mu + F'_\infty \tilde{\gamma}^\perp(Y) \\ &= \int_X F(\tilde{\sigma} \circ f) \, d\mu + F'_\infty \gamma^\perp(X) = D_F(\gamma || \mu). \end{aligned}$$

□

2 Entropy-Transport problem and distances

Let $\gamma \in \mathcal{M}(X \times X)$. In the sequel we denote by $\gamma_i := (\pi^i)_\# \gamma$ the marginals of γ . We are now ready to define the Entropy-Transport problem.

Definition 1 Let $F \in \Gamma_0(\mathbb{R}_+)$ and let $\mathbf{c} : X \times X \rightarrow [0, +\infty]$ be a lower semicontinuous function. The Entropy-Transport functional between the measures $\mu_1, \mu_2 \in \mathcal{M}(X)$ is the functional

$$\begin{aligned} \mathcal{ET}(\cdot || \mu_1, \mu_2) : \mathcal{M}(X \times X) &\rightarrow [0, +\infty], \\ \mathcal{ET}(\gamma || \mu_1, \mu_2) &:= D_F(\gamma_1 || \mu_1) + D_F(\gamma_2 || \mu_2) + \int_{X \times X} \mathbf{c}(x_1, x_2) d\gamma(x_1, x_2). \end{aligned} \quad (25)$$

We define the Entropy-Transport problem between μ_1 and μ_2 as the minimization problem

$$\text{ET}(\mu_1, \mu_2) := \inf_{\gamma \in \mathcal{M}(X \times X)} \mathcal{ET}(\gamma || \mu_1, \mu_2). \quad (26)$$

To highlight the role of the entropy function F and the cost function \mathbf{c} , we also say that ET is the cost of the Entropy-Transport problem induced by (F, \mathbf{c}) .

We are particularly interested in cost functions of the form $\mathbf{c}(x_1, x_2) = \ell(d(x_1, x_2))$ for a certain function $\ell : [0, \infty) \rightarrow [0, \infty]$.

In the next Proposition we recall some properties of Entropy-Transport problems (for a proof see [29]).

Proposition 1 *Let us suppose that the Entropy-Transport problem between the measures $\mu_1, \mu_2 \in \mathcal{M}(X)$ is feasible, i.e. there exists $\gamma \in \mathcal{M}(X \times X)$ such that $\mathcal{ET}(\gamma || \mu_1, \mu_2) < \infty$, and that F is superlinear. Then the infimum in (26) can be replaced by a minimum and the set of minimizers is a compact convex subset of $\mathcal{M}(X \times X)$. Moreover, the functional ET is convex and positively 1-homogeneous (thus subadditive).*

Remark 1 An important role in the theory of Entropy-Transport problems is played by the *marginal perspective cost* H , that we are going to define.

Given a number $c \in [0, +\infty)$ and an admissible entropy function F , we first introduce the marginal perspective function $H_c : [0, +\infty) \times [0, +\infty) \rightarrow [0, +\infty]$ as the lower semicontinuous envelope of the function

$$\tilde{H}_c(r_1, r_2) := \inf_{\theta > 0} R\left(\frac{r_1}{\theta}\right)\theta + R\left(\frac{r_2}{\theta}\right)\theta + \theta c,$$

where R is the reverse entropy defined in (20). If $c = +\infty$, we set

$$H_\infty(r_1, r_2) = F(0)r_1 + F(0)r_2.$$

When $\mathbf{c} : X_1 \times X_2 \rightarrow [0, +\infty]$ is a lower semicontinuous cost function on two metric spaces X_1, X_2 , the induced marginal perspective cost

$$H : X_1 \times [0, +\infty) \times X_2 \times [0, +\infty) \rightarrow [0, +\infty]$$

is defined as

$$H(x_1, r; x_2, t) := H_{\mathbf{c}(x_1, x_2)}(r, t). \quad (27)$$

One can give some equivalent formulations of the problem (26) in terms of the marginal perspective cost (see for instance [29, Theorem 5.8]). Moreover, the metric properties of the entropy-transport cost ET defined in (26) can be read in terms of the properties of H , studied as a function on the space $\mathcal{C}(X) \times \mathcal{C}(X)$. This point of view, which links the Entropy-Transport structure with the conical geometry of the problem, has been deeply investigated by Liero, Mielke and Savaré for the Hellinger-Kantorovich distance [29, Sect. 7] (see also [10, 14] and [15, Chapters 3,4] for general marginal perspective functions).

For brevity, we do not enter into the details of all these formulations (but see Sect. 5 for some details on the conical construction performed in [39]). Here we only remark that for any complete and separable metric space (X, \mathbf{d}) the cost ET^a induces a distance on the space of measures $\mathcal{M}(X)$ if and only if H^a is a distance on the cone $\mathcal{C}(X)$, $a \in (0, 1]$. In general, it is not difficult to identify conditions on F and \mathbf{c} for which the induced function H is nonnegative, symmetric and $H(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) = (x_2, t)$ as points on the cone (see [14, Proposition 4]); on the contrary, proving the triangle inequality for (a power of) H is a much more challenging problem.

2.1 Regular Entropy-Transport distances

In the next definition we introduced the class of *regular Entropy-Transport distances*.

Definition 2 We say that D_{ET} is a regular Entropy-Transport distance if

- There exist $a \in (0, 1]$, $F \in \Gamma_0(\mathbb{R}_+)$ and a function $\ell : [0, \infty) \rightarrow [0, \infty]$ such that for every complete and separable metric space (X, \mathbf{d}) , setting $\mathbf{c}(x_1, x_2) := \ell(\mathbf{d}(x_1, x_2))$, the function D_{ET} coincides with the power a of the Entropy-Transport cost ET induced by (F, \mathbf{c}) , namely

$$\text{D}_{\text{ET}}(\mu_1, \mu_2) = \text{ET}^a(\mu_1, \mu_2) \quad \text{for every } \mu_1, \mu_2 \in \mathcal{M}(X). \quad (28)$$

- The function ℓ is continuous, convex and $\ell(s) = 0$ if and only if $s = 0$.
- F is superlinear and finite valued.
- For every complete and separable metric space (X, \mathbf{d}) , the related Entropy-Transport distance D_{ET} is a complete and separable metric on $\mathcal{M}(X)$ inducing the weak topology.

We also write that the distance D_{ET} is induced by (a, F, ℓ) with obvious meaning.

We notice that if D_{ET} is a regular Entropy-Transport distance induced by (a, F, ℓ) then ℓ is an increasing function and $\lim_{d \rightarrow +\infty} \ell(d) = +\infty$.

We conclude the section with a list of examples of regular Entropy-Transport distances.

Examples 1 (1) Hellinger-Kantorovich: Let $F(s) = U_1(s) := s \log s - s + 1$ and

$$\ell_{\text{HK}}(d) := \begin{cases} -\log(\cos^2(d)) & \text{if } d < \frac{\pi}{2}, \\ +\infty & \text{otherwise.} \end{cases}$$

It is proved in [29, Sect. 7] that $(1/2, U_1, \ell_{\text{HK}})$ induces a regular Entropy-Transport distance, called *Hellinger-Kantorovich* distance. We refer also to [28] for a discussion on “weighted versions” of the Hellinger-Kantorovich distance.

(2) **Gaussian Hellinger-Kantorovich:** Let $F(s) = U_1(s) = s \log s - s + 1$ and $\ell_2(d) := d^2$.

The triple $(1/2, U_1, \ell_2)$ induces a regular Entropy-Transport distance, as discussed in [29, Sect. 7.8]. It is called *Gaussian Hellinger-Kantorovich* distance.

(3) **Quadratic power-like distances:** Let

$$F(s) = U_p(s) := \frac{s^p - p(s-1) - 1}{p(p-1)}, \quad p > 1$$

and $\ell_2(d) = d^2$.

Then, for every $1 < p \leq 3$ the triple $(1/2, U_p, \ell_2)$ induces a regular Entropy-Transport distance, as proved in [14, Theorem 6 and Corollary 1].

We notice that the class of entropy functions $\{U_p\}$ satisfies $\lim_{p \rightarrow 1} U_p(s) = U_1(s)$, justifying the notation we have used (see also [29, Example 2.5]).

(4) **Linear power-like distances:** Let

$$F(s) = U_p(s) := \frac{s^p - p(s-1) - 1}{p(p-1)}, \quad p > 1$$

and $\ell_1(d) := d$.

For every $p > 1$, $(1/2, U_p, \ell_1)$ induces a regular Entropy-Transport distance (see again [14, Theorem 6 and Corollary 1]).

3 Sturm-Entropy-Transport distance

We say that two metric measure spaces (X_1, d_1, μ_1) and (X_2, d_2, μ_2) are *isomorphic* if there exists an isometry $\psi : \text{supp}(\mu_1) \rightarrow \text{supp}(\mu_2)$ such that $\psi_{\#} \mu_1 = \mu_2$, where

$\psi_{\#}$ denotes the push-forward through the map ψ . A necessary condition in order to be isomorphic is that $\mu_1(X_1) = \mu_2(X_2)$.

The family of all isomorphism classes of metric measure spaces will be denoted by \mathbf{X} . From now on, we will identify a metric measure space with its class.

We recall now the definition of the \mathbf{D}_p -distance due to Sturm.

Definition 3 ([43]) Fix $p \geq 1$. Let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be two metric measure spaces, the Sturm \mathbf{D}_p -distance is defined as

$$\mathbf{D}_p((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) := \inf \mathcal{W}_p(\psi_{\#}^1 \mu_1, \psi_{\#}^2 \mu_2), \quad (29)$$

where the infimum is taken over all complete and separable metric spaces (\hat{X}, \hat{d}) with isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$ and $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$.

It is proved in [43, Theorem 3.6] that \mathbf{D}_p is a complete, separable and geodesic metric on the set

$$\mathbf{X}_{1,p} := \{(X, d, \mu) \in \mathbf{X} : \mu \in \mathcal{P}_p(X, d)\}.$$

We are now going to define the *Sturm-Entropy-Transport distance* in a similar way.

Definition 4 Let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be two metric measure spaces, we define the Sturm-Entropy-Transport distance induced by the regular Entropy-Transport distance \mathbf{D}_{ET} as

$$\mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) := \inf \mathbf{D}_{\text{ET}}(\psi_{\#}^1 \mu_1, \psi_{\#}^2 \mu_2), \quad (30)$$

where the infimum is taken over all complete and separable metric spaces (\hat{X}, \hat{d}) with isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$ and $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$.

It is not difficult to prove that the definition is well-posed. Indeed, let us suppose (X'_i, d'_i, μ'_i) is isomorphic to (X_i, d_i, μ_i) through the map φ^i , $i = 1, 2$. Then, for every metric space \hat{X} and every isometric embedding $\psi^i : \text{supp}(\mu_i) \rightarrow \hat{X}$, $i = 1, 2$, we have that

$$\mathbf{D}_{\text{ET}}((\psi^1 \circ \varphi^1)_{\#} \mu_1, (\psi^2 \circ \varphi^2)_{\#} \mu_2) = \mathbf{D}_{\text{ET}}(\psi_{\#}^1 \mu_1, \psi_{\#}^2 \mu_2).$$

It is often convenient to work with explicit realisations of the ambient space (\hat{X}, \hat{d}) , a particularly useful one is given by the disjoint union that we now discuss.

Given two metric spaces (X_1, d_1, μ_1) and (X_2, d_2, μ_2) , let $X_1 \sqcup X_2$ be their disjoint union. We say that a (resp. pseudo-)metric \hat{d} on $X_1 \sqcup X_2$ is a (resp. pseudo-)metric coupling between d_1 and d_2 if $\hat{d}(x, y) = d_1(x, y)$ when $x, y \in X_1$ and $\hat{d}(x, y) = d_2(x, y)$ when $x, y \in X_2$.

A finite valued metric coupling \hat{d} between d_1 and d_2 always exists: to construct it, fix two points $\bar{x}_1 \in X_1, \bar{x}_2 \in X_2$, a number $c \in \mathbb{R}_+$, and define \hat{d} as:

$$\hat{d}(x, y) := \begin{cases} d_1(x, y) & \text{if } x, y \in X_1 \\ d_2(x, y) & \text{if } x, y \in X_2 \\ d_1(x, \bar{x}_1) + c + d_2(\bar{x}_2, y) & \text{if } x \in X_1, y \in X_2 \\ d_1(y, \bar{x}_1) + c + d_2(\bar{x}_2, x) & \text{if } y \in X_1, x \in X_2. \end{cases} \quad (31)$$

Moreover, from any finite valued pseudo-metric coupling \hat{d} of d_1 and d_2 and any $\delta > 0$ we can obtain a complete, separable metric \hat{d}_δ which is again a coupling of d_1 and d_2 in the following way:

$$\hat{d}_\delta := \begin{cases} \hat{d} & \text{on } (X_1 \times X_1) \sqcup (X_2 \times X_2) \\ \hat{d} + \delta & \text{on } (X_1 \times X_2) \sqcup (X_2 \times X_1). \end{cases} \quad (32)$$

We say that a measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ is a *measure coupling* between μ_1 and μ_2 if

$$\gamma(A \times X_2) = \mu_1(A) \text{ and } \gamma(X_1 \times B) = \mu_2(B), \quad (33)$$

for all Borel sets $A \subset X_1$ and $B \subset X_2$. We keep the notation γ_i for the marginals of the measure $\gamma \in \mathcal{M}(X_1 \times X_2)$, $i = 1, 2$.

A more explicit formulation of the function \mathbf{D}_{ET} is given in the following Proposition.

Proposition 2 *Let (X_1, d_1, μ_1) and (X_2, d_2, μ_2) be two metric measure spaces and \mathbf{D}_{ET} a regular Entropy-Transport distance induced by (a, F, ℓ) .*

- (i) *In Definition 4 we can suppose without loss of generality that $\hat{X} = X_1 \sqcup X_2$, $\psi^1 = \iota_1$, $\psi^2 = \iota_2$ be respectively the inclusion of X_1 and X_2 in $X_1 \sqcup X_2$ and the infimum is taken over all the pseudo-metric couplings \hat{d} between d_1 and d_2 .*
- (ii) *In the situation of (i) we will identify μ_k with $(\iota_k)_\# \mu_k$, $k = 1, 2$, and it holds*

$$\begin{aligned} & \mathbf{D}_{\text{ET}}^{1/a}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) \\ &= \inf_C \left\{ \sum_{i=1}^2 D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{d}(x, y)) d\gamma \right\}, \end{aligned} \quad (34)$$

where

$$\begin{aligned} C := \{(\gamma, \hat{d}) : \gamma \in \mathcal{M}(X_1 \times X_2), \\ \hat{d} \text{ finite valued pseudo-metric coupling for } d_1, d_2\} \end{aligned} \quad (35)$$

Proof (i) We first show that the infimum as in (i) is less or equal to the infimum as in Definition 4. Let (\hat{X}, \hat{d}) be a complete and separable metric space with isometric

embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$, $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$, and let $\hat{\gamma} \in \mathcal{M}(\hat{X} \times \hat{X})$. It is immediate to check that

$$\tilde{d}(x_1, x_2) := \begin{cases} d_1(x_1, x_2) & \text{if } (x_1, x_2) \in X_1 \times X_1 \\ d_2(x_1, x_2) & \text{if } (x_1, x_2) \in X_2 \times X_2 \\ \inf_{\substack{y_1 \in \text{supp}(\mu_1) \\ y_2 \in \text{supp}(\mu_2)}} d_1(x_1, y_1) + \hat{d}(\psi^1(y_1), \psi^2(y_2)) + d_2(y_2, x_2) & \text{if } (x_1, x_2) \in X_1 \times X_2 \\ \inf_{\substack{y_1 \in \text{supp}(\mu_1) \\ y_2 \in \text{supp}(\mu_2)}} d_1(x_2, y_1) + \hat{d}(\psi^1(y_1), \psi^2(y_2)) + d_2(y_2, x_1) & \text{if } (x_1, x_2) \in X_2 \times X_1 \end{cases} \quad (36)$$

defines a pseudo-metric on $X_1 \sqcup X_2$, coupling between d_1 and d_2 .

Moreover, setting the Borel injective functions $\Psi^i : \psi^1(\text{supp}(\mu_1)) \cup \psi^2(\text{supp}(\mu_2)) \subset \hat{X} \rightarrow X_1 \sqcup X_2$, $i = 1, 2$, defined as

$$\begin{aligned} \Psi^1(\hat{x}) &:= \begin{cases} \iota_1((\psi^1)^{-1}(\hat{x})) & \text{if } \hat{x} \in \psi^1(\text{supp}(\mu_1)) \\ \iota_2((\psi^2)^{-1}(\hat{x})) & \text{if } \hat{x} \in \psi^2(\text{supp}(\mu_2)), \hat{x} \notin \psi^1(\text{supp}(\mu_1)), \end{cases} \\ \Psi^2(\hat{x}) &:= \begin{cases} \iota_1((\psi^1)^{-1}(\hat{x})) & \text{if } \hat{x} \in \psi^1(\text{supp}(\mu_2)), \hat{x} \notin \psi^2(\text{supp}(\mu_1)) \\ \iota_2((\psi^2)^{-1}(\hat{x})) & \text{if } \hat{x} \in \psi^2(\text{supp}(\mu_2)), \end{cases} \end{aligned}$$

and using Lemma 4 it is immediate to check that $\tilde{\gamma} := (\Psi^1, \Psi^2)_\# \hat{\gamma} \in \mathcal{M}((X_1 \sqcup X_2) \times (X_1 \sqcup X_2))$ satisfies

$$\begin{aligned} & \sum_{i=1}^2 D_F(\tilde{\gamma}_i || (\iota_i)_\# \mu_i) + \int_{\iota_1(X_1) \times \iota_2(X_2)} \ell(\tilde{d}(x, y)) d\tilde{\gamma}(x, y) \\ & \leq \sum_{i=1}^2 D_F(\hat{\gamma}_i || (\psi^i)_\# \mu_i) + \int_{\psi^1(\text{supp}(\mu_1)) \times \psi^2(\text{supp}(\mu_2))} \ell(\hat{d}(x, y)) d\hat{\gamma}(x, y) \\ & \leq \sum_{i=1}^2 D_F(\hat{\gamma}_i || (\psi^i)_\# \mu_i) + \int_{\hat{X} \times \hat{X}} \ell(\hat{d}(x, y)) d\hat{\gamma}(x, y), \end{aligned} \quad (37)$$

where we have used the fact that $\tilde{d}(\Psi^1(x), \Psi^2(y)) \leq \hat{d}(x, y)$ whenever $x \in \psi^1(\text{supp}(\mu_1))$ and $y \in \psi^2(\text{supp}(\mu_2))$.

This yields that the infimum as in (i) is less or equal to the infimum as in Definition 4.

To show that the infimum as in Definition 4 is less or equal to the infimum as in (i), it is sufficient to notice that for every pseudo-metric coupling \hat{d} , for every measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ and for every $\epsilon > 0$ there is $\delta > 0$ such that the complete and separable metric \hat{d}_δ defined in (32) is a coupling between d_1 and d_2 satisfying

$$\int_{X_1 \times X_2} \ell(\hat{d}_\delta(x, y)) d\gamma \leq \int_{X_1 \times X_2} \ell(\hat{d}(x, y)) d\gamma + \epsilon, \quad (38)$$

as a consequence of the finiteness of the measure γ and the continuity of ℓ .

(ii) In case the infimum runs over the couples $(\gamma, \hat{d}) \in C$ such that \hat{d} is a complete and separable metric, the inequality “ \leq ” in (34) is a simple consequence of the explicit formulation of the Entropy-Transport problem together with the fact that the superlinearity of F allows to consider measures $\gamma \in \mathcal{M}((X_1 \sqcup X_2) \times (X_1 \sqcup X_2))$ with support contained in $X_1 \times X_2$. The fact that “ \leq ” holds in (34) even if the infimum is taken over the larger set C is a consequence of (38).

The proof of the inequality “ \geq ” in (34) is analogous to the first part of the proof of (i), see in particular (37). \square

In the next Lemma we collect some of the basic properties of the function \mathbf{D}_{ET} .

Lemma 5 *Let D_{ET} be a regular Entropy-Transport distance induced by (a, F, ℓ) .*

(i) *For any $M \geq 0$ it holds*

$$\mathbf{D}_{\text{ET}}((X_1, d_1, M\mu_1), (X_2, d_2, M\mu_2)) = M^a \mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)). \quad (39)$$

(ii) *If $(X_1, d_1) = (X_2, d_2)$ then*

$$\mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) \leq D_{\text{ET}}(\mu_1, \mu_2). \quad (40)$$

(iii) *The set*

$$\mathbf{X}_* := \left\{ (X, d, \mu) \in \mathbf{X}, \text{supp}(\mu) = \{x_1, \dots, x_n\}, n \in \mathbb{N}, \mu = M \sum_{i=1}^n \delta_{x_i}, M \in \mathbb{R}_+ \right\} \quad (41)$$

is dense in $(\mathbf{X}, \mathbf{D}_{\text{ET}})$.

(iv) *If*

$$\mu = M \sum_{i=1}^n \delta_{x_i} \quad \text{and} \quad \mu' = M \sum_{i=1}^n \delta_{x'_i}, \quad (42)$$

then

$$\mathbf{D}_{\text{ET}}^{1/a}((X, d, \mu), (X', d', \mu')) \leq Mn \ell \left(\sup_{i,j} |d_{ij} - d'_{ij}| \right), \quad (43)$$

where we put $d_{ij} = d(x_i, x_j)$ and $d'_{ij} = d(x'_i, x'_j)$.

(v) *For any $N > 1$ there exists a constant C such that for every M , $1/N < M < N$, we have*

$$\mathbf{D}_{\text{ET}}^{1/a}((X, d, \mu), (X, d, M\mu)) \leq C\mu(X)|M - 1|. \quad (44)$$

Proof (i) This is a consequence of the 1-homogeneity of the cost ET (Proposition 1) and of the push-forward map together with the definitions of D_{ET} and \mathbf{D}_{ET} .

(ii) The result follows from the definition of \mathbf{D}_{ET} , since $(\hat{X}, \hat{d}) = (X_1, d_1)$ with $\psi_1 = \psi_2 = \text{Id}$ is an admissible competitor for the infimum.

- (iii) The result follows by the point (ii) of the present Lemma, the fact that D_{ET} metrizes the weak convergence and the density in $\mathcal{M}(X)$ of the measures μ of the form $M \sum_{i=1}^n \delta_{x_i}$ with respect to weak convergence.
- (iv) Let assume without loss of generality that $X = \{x_1, \dots, x_n\}$ and $X' = \{x'_1, \dots, x'_n\}$. We put $\delta = \sup_{i,j} |d_{ij} - d'_{ij}|$. We construct the following pseudo-metric coupling: on $X \times X$ we define $\hat{d} = d$, on $X' \times X'$ we put $\hat{d} = d'$, on $X \times X'$ we define

$$\hat{d}(x_i, x'_j) := \inf_{k \in \{1, \dots, n\}} d(x_i, x_k) + d'(x'_k, x'_j) + \delta,$$

finally on $X' \times X$ we put

$$\hat{d}(x'_i, x_j) := \inf_{k \in \{1, \dots, n\}} d(x_j, x_k) + d'(x'_k, x'_i) + \delta,$$

so that $\hat{d}(x_i, x'_i) = \hat{d}(x'_i, x_i) = \delta$.

We then define the measure coupling

$$\gamma = M \sum_{i=1}^n \delta_{(x_i, x'_i)}.$$

It is straightforward to see that \hat{d} and γ are actually couplings between d, d' and μ, μ' , respectively. Then, using Proposition 2 and recalling that ℓ is an increasing function we have that

$$D_{\text{ET}}^{1/a}((X, d, \mu), (X', d', \mu')) \leq \int_{X \times X'} \ell(\delta) d\gamma = Mn\ell(\delta),$$

and the thesis follows.

- (v) We can take d itself as metric coupling. Then, by the point (ii) of the present Lemma, we have

$$D_{\text{ET}}^{1/a}((X, d, \mu), (X, d, M\mu)) \leq D_{\text{ET}}^{1/a}(\mu, M\mu) = \text{ET}(\mu, M\mu).$$

By replacing the cost c with the cost

$$c_\infty(x_1, x_2) := \begin{cases} 0 & \text{if } x_1 = x_2 \\ +\infty & \text{otherwise,} \end{cases}$$

we obtain that

$$\text{ET}(\mu, M\mu) \leq \text{ET}_\infty(\mu, M\mu),$$

where we have denoted by ET_∞ the Entropy-Transport problem induced by the entropy function F and the cost \mathbf{c}_∞ . Observe that every admissible entropy function satisfies

$$F(s) \leq C|s - 1|, \quad \text{for every } 1/N < s < N, \quad (45)$$

where

$$C := \max \left\{ \frac{F(1/N)}{1/N - 1}, \frac{F(N)}{N - 1} \right\}.$$

The conclusion now follows from an explicit computation of ET_∞ together with the bound (45). Indeed, we have (see [29, Example E.5])

$$\begin{aligned} \text{ET}_\infty(\mu, M\mu) &\leq \min_{\theta \in [1, M]} \int_X C|\theta - 1| + CM|\theta/M - 1| d\mu \\ &= C\mu(X)|M - 1|. \end{aligned} \quad (46)$$

□

The next Lemma shows the existence of the optimal couplings.

Lemma 6 *Let D_{ET} be a regular Entropy-Transport distance induced by (a, F, ℓ) . Let $(X_1, \mathbf{d}_1, \mu_1)$ and $(X_2, \mathbf{d}_2, \mu_2)$ be two metric measure spaces. Then:*

- (i) *here exist a measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ and a pseudo-metric coupling $\hat{\mathbf{d}}$ between \mathbf{d}_1 and \mathbf{d}_2 such that*

$$\begin{aligned} \mathbf{D}_{\text{ET}}^{1/a}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ = \sum_{i=1}^2 D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{\mathbf{d}}(x, y)) d\gamma. \end{aligned} \quad (47)$$

- (ii) *There exist a complete and separable metric space $(\tilde{X}, \tilde{\mathbf{d}})$ and isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \tilde{X}$, $\psi^2 : \text{supp}(\mu_2) \rightarrow \tilde{X}$ such that*

$$\mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = (D_{\text{ET}})_{\tilde{\mathbf{d}}}(\psi_{\#}^1 \mu_1, \psi_{\#}^2 \mu_2), \quad (48)$$

where we have denoted by $(D_{\text{ET}})_{\tilde{\mathbf{d}}}$ the Entropy-Transport distance computed in the space $(\tilde{X}, \tilde{\mathbf{d}})$.

Proof (i) **Step 1:** tightness of the plans.

By Proposition 2 there exist a sequence $\gamma_n \in \mathcal{M}(X_1 \times X_2)$ and $\hat{\mathbf{d}}_n$ pseudo-

metric couplings of d_1, d_2 such that

$$\sum_{i=1}^2 D_F((\gamma_n)_i || \mu_i) + \int_{X_1 \times X_2} \ell \left(\hat{d}_n(x, y) \right) d\gamma_n < D_{\text{ET}}^{1/a}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) + \frac{1}{n}. \quad (49)$$

Since the entropy functionals with respect to the fixed measures μ_1 and μ_2 are bounded, we can apply Theorems 1 and Lemma 3 in order to obtain the existence of subsequences (from now on we will not relabel them) such that $(\gamma_n)_i$ converges weakly to some $\gamma^i \in \mathcal{M}(X_i)$, $i = 1, 2$. Since $(\gamma_n)_i$ are marginals of the measure γ_n , the tightness of $(\gamma_n)_i$ implies the tightness of γ_n , so that the sequence $\gamma_n \in \mathcal{M}(X_1 \times X_2)$ is converging to some γ . Moreover, by the continuity of the operator π_{\sharp}^i with respect to the weak topology, the marginals of γ coincide with γ^i , $i = 1, 2$. We notice that if γ is the null measure the proof is concluded by taking any pseudo-metric coupling \hat{d} between d_1 and d_2 .

Step 2: pre-compactness of the pseudo-metric couplings.

Regarding the sequence \hat{d}_n , by the triangle inequality we have that

$$|\hat{d}_n(x_1, y_1) - \hat{d}_n(x_2, y_2)| \leq |d_1(x_1, x_2) + d_2(y_1, y_2)|.$$

In particular, \hat{d}_n is uniformly 1-Lipschitz with respect to the complete and separable metric $d_1 + d_2$ on $X_1 \times X_2$. We claim it is also uniformly bounded in a point. To see this, take $(\bar{x}, \bar{y}) \in \text{supp}(\gamma)$: since γ_n weakly converges to γ for every $r, \epsilon > 0$ and for all n sufficiently large we have

$$\gamma_n(B_r(\bar{x}) \times B_r(\bar{y})) \geq \gamma(B_r(\bar{x}) \times B_r(\bar{y})) - \epsilon.$$

Fix $r > 0$ and suppose by contradiction that there exists a subsequence (not relabeled) such that $2r \leq \hat{d}_n(\bar{x}, \bar{y}) \rightarrow +\infty$. For $\epsilon = \epsilon(r)$ small enough, from (49), the fact that $(\bar{x}, \bar{y}) \in \text{supp}(\gamma)$ and ℓ is increasing we infer the existence of some positive constants C, c such that for all n sufficiently large

$$\begin{aligned} C &> \int_{X_1 \times X_2} \ell \left(\hat{d}_n(x, y) \right) d\gamma_n(x, y) \\ &\geq \int_{B_r(\bar{x}) \times B_r(\bar{y})} \ell \left(\hat{d}_n(\bar{x}, \bar{y}) - 2r \right) d\gamma_n(x, y) \\ &\geq \ell \left(\hat{d}_n(\bar{x}, \bar{y}) - 2r \right) [\gamma(B_r(\bar{x}) \times B_r(\bar{y})) - \epsilon] \geq c\ell \left(\hat{d}_n(\bar{x}, \bar{y}) - 2r \right). \end{aligned}$$

Since ℓ has bounded sublevels, this implies that there exists a constant K such that $\hat{d}_n(\bar{x}, \bar{y}) < K$ for every n that leads to a contradiction.

We can thus apply Ascoli-Arzelà's theorem to infer the existence of a limit function $d : X_1 \times X_2 \rightarrow [0, \infty)$ such that d_n converges (up to subsequence) pointwise to d and the convergence is uniform on compact sets. We can extend d to $(X_1 \sqcup X_2) \times (X_1 \sqcup X_2)$ in order to get a limit pseudo-metric coupling, that we denote in the same way.

Step 3: passing to the limit.

Next, we pass to the limit in the following expression

$$\sum_{i=1}^2 D_F((\gamma_n)_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{d}_n(x, y)) d\gamma_n.$$

By Lemma 2, the entropy is jointly lower semicontinuous and thus

$$\liminf_n D_F((\gamma_n)_i || \mu_i) \geq D_F(\gamma_i || \mu_i).$$

So, it is sufficient to prove that

$$\liminf_n \int_{X_1 \times X_2} \ell(\hat{d}_n(x, y)) d\gamma_n \geq \int_{X_1 \times X_2} \ell(\hat{d}(x, y)) d\gamma. \quad (50)$$

Using the equi-tightness of $\{\gamma_k\}$ we can find a sequence of compact sets $K_{1,n} \subset X_1$ and $K_{2,n} \subset X_2$ such that

$$\gamma_k(X_1 \times X_2 \setminus (K_{1,n} \times K_{2,n})) \leq \frac{1}{n}$$

for every k . We define $\ell_m(r) := \min(\ell(r), m)$, so that the sequence of functions $(x, y) \mapsto \ell_m(d_n(x, y))$ converges uniformly on compact subsets of $X_1 \times X_2$, as $n \rightarrow \infty$. Possibly by taking a further subsequence via a diagonal argument, we can infer that $\|\ell_m(d) - \ell_m(d_n)\|_{\infty;n} \rightarrow 0$ when $n \rightarrow \infty$, where we denote by $\|\cdot\|_{\infty;n}$ the supremum norm in the set $K_{1,n} \times K_{2,n}$. Let M be a positive constant such that $\gamma_n(X_1 \times X_2) \leq M$ for every n . We can bound the integral on the left hand side of (50) in the following way:

$$\begin{aligned} \int_{X_1 \times X_2} \ell(\hat{d}_n) d\gamma_n &\geq \int_{X_1 \times X_2} \ell_m(\hat{d}_n) d\gamma_n \geq \int_{K_{1,n}^1 \times K_{2,n}^2} \ell_m(\hat{d}_n) d\gamma_n \\ &\geq \int_{K_{1,n}^1 \times K_{2,n}^2} \ell_m(\hat{d}) d\gamma_n - M \|\ell_m(\hat{d}) - \ell_m(\hat{d}_n)\|_{\infty;n} \\ &\geq \int_{X_1 \times X_2} \ell_m(\hat{d}) d\gamma_n - M \|\ell_m(\hat{d}) - \ell_m(\hat{d}_n)\|_{\infty;n} - m/n. \end{aligned}$$

Now we can pass to the limit with respect to n using the weak convergence of $\{\gamma_n\}$, and we obtain

$$\liminf_n \int_{X_1 \times X_2} \ell(\hat{\mathbf{d}}_n) d\gamma_n \geq \int_{X_1 \times X_2} \ell_m(\hat{\mathbf{d}}) d\gamma$$

and then we conclude using the Beppo Levi's monotone convergence theorem with respect to m .

- (ii) Without loss of generality we assume $\text{supp}(\mu_i) = X_i$. By the previous point we know the existence of an optimal measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ and an optimal pseudo-metric coupling $\hat{\mathbf{d}}$ between \mathbf{d}_1 and \mathbf{d}_2 . We consider the complete and separable metric space $(\tilde{X}, \tilde{\mathbf{d}})$ constructed as in Lemma 1. Denoting by $p : X_1 \sqcup X_2 \rightarrow \tilde{X}$ the projection to the quotient and using the identification

$$X_1 \sqcup X_2 = X_1 \times \{0\} \cup X_2 \times \{1\},$$

we notice that $X_1 \times X_2 \hookrightarrow \tilde{X} \times \tilde{X}$ via the injective Borel map

$$\psi(x_1, x_2) = (\psi^1(x_1), \psi^2(x_2)) := (p(x_1, 0), p(x_2, 1)).$$

Moreover, we also have that ψ^i is an isometry of (X_i, \mathbf{d}_i) onto its image in $(\tilde{X}, \tilde{\mathbf{d}})$, $i = 1, 2$. Thus, denoting by γ_i the marginals of γ , we can consider the measures $\psi_{\#}\gamma$ whose projections are $(\psi^1)_{\#}\gamma_1$ and $(\psi^2)_{\#}\gamma_2$. Using Lemma 4 we know that

$$D_F(\gamma_i || \mu_i) = D_F((\psi^i)_{\#}\gamma_i || (\psi^i)_{\#}\mu_i), \quad i = 1, 2. \quad (51)$$

By recalling the definition of $\tilde{\mathbf{d}}$, we also have

$$\int_{X_1 \times X_2} \ell(\hat{\mathbf{d}}(x, y)) d\gamma = \int_{\tilde{X} \times \tilde{X}} \ell(\tilde{\mathbf{d}}(x, y)) d(\psi_{\#}\gamma). \quad (52)$$

Thus, as a consequence of (51), (52) and the optimality of γ and $\hat{\mathbf{d}}$, the equality (48) holds on $(\tilde{X}, \tilde{\mathbf{d}})$ (with optimal measure $\psi_{\#}\gamma$).

□

Remark 2 It is clear that the optimal coupling $\hat{\mathbf{d}}$ whose existence is proven in the previous Lemma is in general only a pseudo-metric and not a metric on $X_1 \sqcup X_2$. To see this, it is sufficient to consider two isomorphic metric measure spaces $(X_1, \mathbf{d}_1, \mu_1)$, $(X_2, \mathbf{d}_2, \mu_2)$. If we denote by $\psi : X_1 \rightarrow X_2$ the isometry between (X_1, \mathbf{d}_1) and (X_2, \mathbf{d}_2) , the optimal coupling $\hat{\mathbf{d}}$ satisfies $\hat{\mathbf{d}}(x_1, \psi(x_1)) = 0$ for μ_1 -a.e. x_1 .

The next theorem is the main result of the paper.

Theorem 2 *Let D_{ET} be a regular Entropy-Transport distance induced by (a, F, ℓ) . Then $(\mathbf{X}, \mathbf{D}_{\text{ET}})$ is a complete and separable metric space. It is also a length (resp. geodesic) space if D_{ET} is a length (resp. geodesic) metric.*

Proof Step 1: \mathbf{D}_{ET} defines a metric.

It is clear that \mathbf{D}_{ET} is symmetric, finite valued, nonnegative and

$$\mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = 0 \text{ if } (X_1, \mathbf{d}_1, \mu_1) = (X_2, \mathbf{d}_2, \mu_2).$$

We claim that $\mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = 0$ implies that the metric measure spaces $(X_1, \mathbf{d}_1, \mu_1)$ and $(X_2, \mathbf{d}_2, \mu_2)$ are isomorphic. By Lemma 6 there exist a measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ and a pseudo-metric coupling $\hat{\mathbf{d}}$ such that

$$0 = \sum_{i=1}^2 D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{\mathbf{d}}(x, y)) d\gamma.$$

All the terms are nonnegative, so that $D_F(\gamma_i || \mu_i) = 0$ and thus $\gamma_i = \mu_i$, $i = 1, 2$. Moreover, since $\ell(d) = 0$ if and only if $d = 0$, it follows that $\hat{\mathbf{d}}(x, y) = 0$ for γ -a.e (x, y) . We also have

$$\hat{\mathbf{d}}(x, y) = 0 \text{ for all } (x, y) \in \text{supp}(\gamma). \quad (53)$$

To see this, let $(\bar{x}, \bar{y}) \in \text{supp}(\gamma)$ so that for every $r > 0$ we have $\gamma(B_r(\bar{x}, \bar{y})) > 0$ where

$$B_r((\bar{x}, \bar{y})) = \{(x, y) \in X_1 \times X_2 \mid \mathbf{d}_1(\bar{x}, x) + \mathbf{d}_2(\bar{y}, y) < r\}.$$

We consider a sequence of balls of radius $r_n := 1/n$, $n \in \mathbb{N}$, and use the fact that $\hat{\mathbf{d}}(x, y) = 0$ for γ -a.e (x, y) to infer the existence of a sequence of points $(x_n, y_n) \in B_{r_n}((\bar{x}, \bar{y}))$ such that $\hat{\mathbf{d}}(x_n, y_n) = 0$. Thus

$$\hat{\mathbf{d}}(\bar{x}, \bar{y}) \leq \mathbf{d}_1(\bar{x}, x_n) + \hat{\mathbf{d}}(x_n, y_n) + \mathbf{d}_2(\bar{y}, y_n) < 1/n.$$

Sending $n \rightarrow +\infty$ and using the arbitrariness of (\bar{x}, \bar{y}) , the claim (53) follows.

Since \mathbf{d}_1 and \mathbf{d}_2 are metrics, we infer that for every $x_1 \in \text{supp}(\mu_1)$ there exists a unique $x_2 \in \text{supp}(\mu_2)$ such that $(x_1, x_2) \in \text{supp}(\gamma)$. Indeed, for any $x_2, \tilde{x}_2 \in \text{supp}(\mu_2)$ such that $(x_1, x_2), (x_1, \tilde{x}_2) \in \text{supp}(\gamma)$ we have

$$\mathbf{d}_2(x_2, \tilde{x}_2) = \hat{\mathbf{d}}(x_2, \tilde{x}_2) \leq \hat{\mathbf{d}}(x_2, x_1) + \hat{\mathbf{d}}(\tilde{x}_2, x_1) = 0$$

and thus $x_2 = \tilde{x}_2$. Switching the role of X_1 and X_2 in the argument above, we obtain the existence of a bijection $\psi : \text{supp}(\mu_1) \rightarrow \text{supp}(\mu_2)$ such that $\gamma = (\text{Id}, \psi)_\# \mu_1$ and (in virtue of (53))

$$\hat{\mathbf{d}}(x, \psi(x)) = 0 \text{ for all } x \in \text{supp}(\mu_1). \quad (54)$$

Let $x, y \in \text{supp}(\mu_1)$, from (54) and the triangle inequality it follows

$$\begin{aligned} d_1(x, y) &= \hat{d}(x, y) \leq \hat{d}(x, \psi(x)) + \hat{d}(\psi(x), \psi(y)) + \hat{d}(y, \psi(y)) \\ &= d_2(\psi(x), \psi(y)), \\ d_2(\psi(x), \psi(y)) &= \hat{d}(\psi(x), \psi(y)) \leq \hat{d}(x, \psi(x)) + \hat{d}(x, y) + \hat{d}(y, \psi(y)) \\ &= d_1(x, y), \end{aligned}$$

which implies that $\psi : \text{supp}(\mu_1) \rightarrow \text{supp}(\mu_2)$ is an isometry.

Hence (X_1, d_1, μ_1) and (X_2, d_2, μ_2) are isomorphic, as claimed.

Regarding the triangle inequality, let (X_i, d_i, μ_i) , $i = 1, 2, 3$, be three metric measure spaces. From the definition of \mathbf{D}_{ET} and Proposition 2, for every $\epsilon > 0$ we find a pseudo-metric coupling d_{12} between d_1 and d_2 , and a pseudo-metric coupling d_{23} between d_2 and d_3 such that

$$\begin{aligned} \mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) &\geq (\mathbf{D}_{\text{ET}})_{d_{12}}(\mu_1, \mu_2) - \epsilon, \\ \mathbf{D}_{\text{ET}}((X_2, d_2, \mu_2), (X_3, d_3, \mu_3)) &\geq (\mathbf{D}_{\text{ET}})_{d_{23}}(\mu_2, \mu_3) - \epsilon, \end{aligned}$$

where we have denoted by $(\mathbf{D}_{\text{ET}})_d$ the Entropy-Transport distance induced by the pseudo-metric d . Set $X := X_1 \sqcup X_2 \sqcup X_3$ and define a pseudo-metric d on X in the following way

$$d(x, y) := \begin{cases} d_{12}(x, y) & \text{if } x, y \in X_1 \sqcup X_2 \\ d_{23}(x, y) & \text{if } x, y \in X_2 \sqcup X_3 \\ \inf_{z \in X_2} [d_{12}(x, z) + d_{23}(z, y)] & \text{if } x \in X_1 \text{ and } y \in X_3 \\ \inf_{z \in X_2} [d_{23}(x, z) + d_{12}(z, y)] & \text{if } x \in X_3 \text{ and } y \in X_1. \end{cases}$$

We notice that d coincides with d_i when restricted to X_i . By applying Proposition 2, the point (ii) of Lemma 5 and the triangle inequality of $(\mathbf{D}_{\text{ET}})_d$ we obtain

$$\begin{aligned} \mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_3, d_3, \mu_3)) &\leq (\mathbf{D}_{\text{ET}})_d(\mu_1, \mu_3) \leq (\mathbf{D}_{\text{ET}})_d(\mu_1, \mu_2) + (\mathbf{D}_{\text{ET}})_d(\mu_2, \mu_3) \\ &= (\mathbf{D}_{\text{ET}})_{d_{12}}(\mu_1, \mu_2) + (\mathbf{D}_{\text{ET}})_{d_{23}}(\mu_2, \mu_3) \\ &\leq \mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) + \mathbf{D}_{\text{ET}}((X_2, d_2, \mu_2), (X_3, d_3, \mu_3)) + 2\epsilon. \end{aligned}$$

The conclusion follows since $\epsilon > 0$ is arbitrary.

Step 2: Completeness of \mathbf{D}_{ET} .

In order to prove completeness, let $\{(X_n, d_n, \mu_n)\}_{n \in \mathbb{N}}$ be a Cauchy sequence in the space $(\mathbf{X}, \mathbf{D}_{\text{ET}})$. In order to have convergence of the full sequence, it is enough to prove that there exists a converging subsequence. Let us consider a subsequence such that

$$\mathbf{D}_{\text{ET}}^{1/a}((X_{n_k}, d_{n_k}, \mu_{n_k}), (X_{n_{k+1}}, d_{n_{k+1}}, \mu_{n_{k+1}})) < 2^{-(k+1)}.$$

By definition of \mathbf{D}_{ET} and Proposition 2, we can find a measure $\gamma_{k+1} \in \mathcal{M}(X_{n_k} \times X_{n_{k+1}})$ and a complete and separable metric coupling $\hat{\mathbf{d}}_{k+1}$ between $\mathbf{d}_{X_{n_k}}$ and $\mathbf{d}_{X_{n_{k+1}}}$ such that

$$\begin{aligned} & \int_{X_{n_k}} F(\sigma_{n_k}) d\mu_{n_k} + \int_{X_{n_{k+1}}} F(\sigma_{n_{k+1}}) d\mu_{n_{k+1}} \\ & + \int_{X_{n_k} \times X_{n_{k+1}}} \ell(\hat{\mathbf{d}}_{k+1}) d\gamma_{k+1} < 2^{-k}, \end{aligned} \quad (55)$$

where σ_{n_k} (resp. $\sigma_{n_{k+1}}$) is the Radon-Nykodim derivative of the first (resp. second) marginal of γ_{k+1} with respect to μ_{n_k} (resp. $\mu_{n_{k+1}}$).

Now we want to define a sequence $\{(X'_k, \mathbf{d}'_k)\}_{k=1}^\infty$ of metric spaces such that $X_{n_k} \subset X'_k$ and $X'_k \subset X'_{k+1}$. We proceed in the following way: we set

$$\begin{aligned} (X'_1, \mathbf{d}'_1) &:= (X_{n_1}, \mathbf{d}_{X_{n_1}}), \\ X'_{k+1} &:= X'_k \sqcup X_{n_{k+1}} / \sim, \end{aligned}$$

where $x \sim y$ if $\mathbf{d}'_{k+1}(x, y) = 0$ and the latter is defined as

$$\mathbf{d}'_{k+1}(x, y) := \begin{cases} \mathbf{d}'_k(x, y) & \text{if } x, y \in X'_k \\ \hat{\mathbf{d}}_{k+1}(x, y) & \text{if } x, y \in X_{n_k} \sqcup X_{n_{k+1}} \\ \inf_{z \in X_{n_k}} \mathbf{d}'_k(x, z) + \hat{\mathbf{d}}_{k+1}(z, y) & \text{if } x \in X'_k, y \in X_{n_{k+1}} \\ \inf_{z \in X_{n_k}} \mathbf{d}'_k(y, z) + \hat{\mathbf{d}}_{k+1}(z, x) & \text{if } y \in X'_k, x \in X_{n_{k+1}}. \end{cases}$$

From the definition of \mathbf{d}'_k , it is clear that we can endow the space $X' := \bigcup_{k=1}^\infty X'_k$ with a limit metric \mathbf{d}' . Now we consider the completion (X, \mathbf{d}) of (X', \mathbf{d}') and we notice that $(X_{n_k}, \mathbf{d}_{X_{n_k}})$ is isometrically embedded in this space for every k . Using the embedding, we can also define a measure $\bar{\mu}_{n_k}$ as the push-forward of the measure μ_{n_k} . Combining the construction above with (55) gives

$$\begin{aligned} (\mathbf{D}_{\text{ET}})_{\mathbf{d}}^{1/a}(\bar{\mu}_{n_k}, \bar{\mu}_{n_{k+1}}) &\leq \int_{X_{n_k}} F(\sigma_{n_k}) d\mu_{n_k} + \int_{X_{n_{k+1}}} F(\sigma_{n_{k+1}}) d\mu_{n_{k+1}} \\ &+ \int_{X_{n_k} \times X_{n_{k+1}}} \ell(\hat{\mathbf{d}}_{k+1}) d\gamma_{k+1} < 2^{-k}, \end{aligned} \quad (56)$$

where $(\mathbf{D}_{\text{ET}})_{\mathbf{d}}$ is the regular Entropy-Transport distance computed in the space (X, \mathbf{d}) . In particular, (56) implies that $(\bar{\mu}_{n_k})_{k \in \mathbb{N}}$ is a Cauchy sequence in $(\mathcal{M}(X), (\mathbf{D}_{\text{ET}})_{\mathbf{d}})$. Since $(\mathbf{D}_{\text{ET}})_{\mathbf{d}}$ is complete, there exists $\mu \in \mathcal{M}(X)$ such that $(\mathbf{D}_{\text{ET}})_{\mathbf{d}}^{1/a}(\bar{\mu}_{n_k}, \mu) \rightarrow 0$.

Using again that $(X_{n_k}, \mathbf{d}_{X_{n_k}})$ is isometrically embedded in (X, \mathbf{d}) and the point (ii) of Lemma 5, we can conclude that

$$\mathbf{D}_{\text{ET}}((X_{n_k}, \mathbf{d}_{n_k}, \mu_{n_k}), (X, \mathbf{d}, \mu)) \leq (\mathbf{D}_{\text{ET}})_{\mathbf{d}}(\bar{\mu}_{n_k}, \mu) \rightarrow 0. \quad (57)$$

Step 3: Separability of \mathbf{D}_{ET} .

Thanks to (iii) of Lemma 5 it is enough to show that the set \mathbf{X}_* , defined in (41), is separable. To this aim, we notice that \mathbf{X}_* can be written as $\bigsqcup_{n \in \mathbb{N}} \tilde{\mathcal{K}}_n$ where

$$\tilde{\mathcal{K}}_n := \{(X, \mathbf{d}, \mu) \in \mathbf{X}_* : \text{supp}(\mu) \text{ has } n \text{ points}\}.$$

Since the set of all $(D, M) = (D_{ij}, M) \in \mathbb{R}_+^{n \times n} \times \mathbb{R}_+$ such that

$$D_{ij} = D_{ji}, \quad D_{ij} = 0 \iff i = j, \quad D_{ij} \leq D_{ik} + D_{kj} \quad (58)$$

is separable (as a subset of the Euclidean space), using (iv) of Lemma 5 we get that

$$\tilde{\mathcal{K}}_{n,M} := \{(X, \mathbf{d}, \mu) \in \mathbf{X}_* : \text{supp}(\mu) \text{ has } n \text{ points and } \mu(X) = nM\}$$

is separable for every fixed $n \in \mathbb{N}$, $M > 0$. The separability of $\tilde{\mathcal{K}}_n$ follows by the separability of $\tilde{\mathcal{K}}_{n,M}$ combined with (v) of Lemma 5.

Step 4: Length/geodesic property of \mathbf{D}_{ET} .

Let us start by proving the length property. Let $(X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2) \in \mathbf{X}$. By definition of \mathbf{D}_{ET} , for every $\varepsilon > 0$ we can find a complete and separable metric space (X, \mathbf{d}) and isometric embeddings $\psi^i : \text{supp}(\mu_i) \rightarrow X$, $i = 1, 2$, such that

$$\mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \geq (\mathbf{D}_{\text{ET}})_d(\mu_1, \mu_2) - \varepsilon, \quad (59)$$

where, as before, we identify $\text{supp}(\mu_i)$ with its isometric image $\psi^i(\text{supp}(\mu_i))$, and correspondingly μ_i with $\psi_{\#}^i \mu_i$, $i = 1, 2$, in order to keep notation short.

Recall that, by slightly modifying the classical Kuratowski embedding, one can show that every complete and separable metric space can be isometrically embedded in a complete, separable and *geodesic* metric space (see for instance [23, Exercise 1c. Ch. 3 $\frac{1}{2}$.1] or [22, Proposition 1.2.12]). Thus, recalling also Lemma 4, without loss of generality we can assume that the complete and separable metric space (X, \mathbf{d}) above is also *geodesic*.

By assumption $(\mathbf{D}_{\text{ET}})_d$ is a length distance on $\mathcal{M}(X)$ since (X, \mathbf{d}) is a length space, so that we can find a curve $(\mu_t)_{t \in [1,2]} \subset (\mathcal{M}(X), (\mathbf{D}_{\text{ET}})_d)$ from μ_1 to μ_2 satisfying

$$\text{Length}_{(\mathbf{D}_{\text{ET}})_d}((\mu_t)_{t \in [1,2]}) \leq (\mathbf{D}_{\text{ET}})_d(\mu_1, \mu_2) + \varepsilon. \quad (60)$$

Now, it is easy to check that the \mathbf{D}_{ET} -length of the curve of m.m.s. $((X, \mathbf{d}, \mu_t))_{t \in [1,2]} \subset \mathbf{X}$ satisfies

$$\text{Length}_{\mathbf{D}_{\text{ET}}}(((X, \mathbf{d}, \mu_t))_{t \in [1,2]}) \leq \text{Length}_{(\mathbf{D}_{\text{ET}})_d}((\mu_t)_{t \in [1,2]}). \quad (61)$$

Indeed the length of a curve is by definition the supremum of the sums of mutual distances over finite partitions (13), and for every partition (t_i) of $[1, 2]$ it holds

$$\begin{aligned} \sum_i \mathbf{D}_{\text{ET}}((X, \mathbf{d}, \mu_{t_{i+1}}), (X, \mathbf{d}, \mu_{t_i})) &\leq \sum_i (\mathbf{D}_{\text{ET}})_d(\mu_{t_{i+1}}, \mu_{t_i}) \\ &\leq \text{Length}_{(\mathbf{D}_{\text{ET}})_d}((\mu_t)_{t \in [1, 2]}). \end{aligned}$$

The combination of (59), (60) and (61) gives

$$\begin{aligned} \text{Length}_{\mathbf{D}_{\text{ET}}}(((X, \mathbf{d}, \mu_t))_{t \in [1, 2]}) &\leq \text{Length}_{(\mathbf{D}_{\text{ET}})_d}((\mu_t)_{t \in [1, 2]}) \leq (\mathbf{D}_{\text{ET}})_d(\mu_1, \mu_2) + \varepsilon \\ &\leq \mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) + 2\varepsilon, \end{aligned}$$

as desired.

To prove the geodesic property in the case \mathbf{D}_{ET} is a geodesic distance, we notice that we can follow verbatim the argument given above with $\varepsilon = 0$. Here one has to notice that the existence of an *optimal* complete and separable metric space on which (59) holds with $\varepsilon = 0$ follows from (ii) of Lemma 6. \square

Remark 3 It is proved in [29, Proposition 8.3] that $(\mathcal{M}(X), \text{HK})$ is a geodesic space when the underlying space (X, \mathbf{d}) is geodesic. In particular, the last claim of Theorem 2 can be applied to the Hellinger-Kantorovich distance.

To the best of our knowledge, up to now this is the only known example of regular Entropy-Transport geodesic distance (with the trivial exception of weighted variants of HK [28]).

3.1 Topology

Let us introduce a notion of convergence for sequences of (equivalence classes of) metric measure spaces (see [21, Definition 3.9] for the corresponding notion in the context of *pointed* metric measure spaces).

Definition 5 We say that a sequence $(X_n, \mathbf{d}_n, \mu_n)_{n \in \mathbb{N}}$ weakly measured-Gromov converges to $(X_\infty, \mathbf{d}_\infty, \mu_\infty)$ if there exist a complete and separable metric space (X, \mathbf{d}) and isometric embeddings $\iota_n : X_n \rightarrow X$, $n \in \mathbb{N}$, such that $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$ weakly in $\mathcal{M}(X)$.

In the next Theorem we see that this notion of convergence actually coincides with the convergence induced by any Sturm-Entropy-Transport distance.

Theorem 3 Let \mathbf{D}_{ET} be a regular Entropy-Transport distance induced by (a, F, ℓ) . A sequence $(X_n, \mathbf{d}_n, \mu_n)_{n \in \mathbb{N}}$ weakly measured Gromov converges to $(X_\infty, \mathbf{d}_\infty, \mu_\infty)$ if and only if

$$\mathbf{D}_{\text{ET}}((X_n, \mathbf{d}_n, \mu_n), (X_\infty, \mathbf{d}_\infty, \mu_\infty)) \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (62)$$

Proof Let us suppose the validity of (62). By definition of \mathbf{D}_{ET} we know that there exist a complete and separable metric space (Y_n, d_{Y_n}) and isometric embeddings ψ_n, ψ_n^∞ of $(X_n, d_n), (X_\infty, d_\infty)$ respectively, in Y_n such that

$$D_{\text{ET}}((\psi_n)_\# \mu_n, (\psi_n^\infty)_\# \mu_\infty) < \frac{1}{n}, \quad (63)$$

where D_{ET} is computed in the space Y_n . We now define $Y := \sqcup_n X_n, n \in \bar{\mathbb{N}}$ endowed with the pseudo-metric d_Y

$$d_Y(y, y') := \begin{cases} d_n(y, y') & \text{if } y, y' \in X_n, n \in \bar{\mathbb{N}} \\ d_{Y_n}(\psi_n(y), \psi_n^\infty(y')) & \text{if } y \in X_n, y' \in X_\infty \\ d_{Y_n}(\psi_n^\infty(y), \psi_n(y')) & \text{if } y \in X_\infty, y' \in X_n \\ \inf_{x \in X_\infty} d_{Y_n}(\psi_n(y), \psi_n^\infty(x)) + d_{Y_m}(\psi_m(y'), \psi_m^\infty(x)) & \text{if } y \in X_n, y' \in X_m. \end{cases}$$

We now consider the space Y / \sim defined as the quotient of Y with respect to the equivalence relation

$$y \sim y' \Leftrightarrow d_Y(y, y') = 0, \quad (64)$$

and we then define the completion of this space, that we still denote by (Y, d_Y) . It is easy to see that Y is separable. By construction we notice that the set

$$\psi_n(X_n) \cup \psi_n^\infty(X_\infty) \subset Y_n$$

endowed with the distance d_{Y_n} is canonically isometrically embedded in (Y, d_Y) , so that every space $X_n, n \in \bar{\mathbb{N}}$, is canonically isometrically embedded into Y by a map ψ'_n . We claim now that Y and ψ'_n provide a realization of the weakly measured Gromov convergence. To see this, it is enough to notice that $(\psi'_n)_\# \mu_n \rightarrow (\psi'_\infty)_\# \mu_\infty$ weakly in $\mathcal{M}(Y)$ which is a consequence of the construction of ψ'_n , (63) and the fact that D_{ET} induces the weak topology.

For the converse, let us suppose that $(X_n, d_n, \mu_n)_{n \in \mathbb{N}}$ weakly measured Gromov converges to $(X_\infty, d_\infty, \mu_\infty)$. By definition we know that there exist a complete and separable metric space (X, d) and isometric embeddings $\iota_n : X_n \rightarrow X, n \in \bar{\mathbb{N}}$, such that $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$ weakly in $\mathcal{M}(X)$. Since D_{ET} metrizes the weak convergence on $\mathcal{M}(X)$ we know that

$$D_{\text{ET}}((\iota_n)_\# \mu_n, (\iota_\infty)_\# \mu_\infty) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and the result follows by the very definition of \mathbf{D}_{ET} , noticing that (X, d) is a possible competitor. \square

Let us denote by $\mathbf{X}(K, N, L, v, V)$ the family of isomorphism classes of metric measure spaces $(X, d, \mu) \in \text{CD}(K, N)$ such that

$$\text{diam}(X) \leq L \quad \text{and} \quad 0 < v \leq \mu(X) \leq V.$$

Let $\tilde{\mathbf{X}}(K, N, L, v, V)$ be the family of isomorphism classes of spaces

$$(X, d, \mu) \in \mathbf{X}(K, N, L, v, V)$$

such that μ has full support.

Theorem 4 Fix $K \in \mathbb{R}$, $N \in (1, \infty)$, $L \in (0, \infty)$ and $0 < v \leq V < \infty$. Let D_{ET} be a regular Entropy-Transport distance. Then

- $\mathbf{X}(K, N, L, v, V)$ is compact with respect to \mathbf{D}_{ET} .
- $\tilde{\mathbf{X}}(K, N, L, v, V)$ is compact with respect to $m\text{GH}$. Moreover on such family the \mathbf{D}_{ET} -topology and the $m\text{GH}$ -topology coincide.

Proof By [21, Corollary 3.22] we have precompactness of $\mathbf{X}(K, N, L, v, V)$ with respect to the weakly measured Gromov convergence and thus precompactness with respect also to the \mathbf{D}_{ET} -convergence by Theorem 3. From [44, Theorem 3.1] (see also [21, Theorem 4.9]) we know that the condition $\text{CD}(K, N)$ is stable with respect to the weakly measured Gromov convergence and thus the first statement follows. For the second statement we observe that the spaces in $\tilde{\mathbf{X}}(K, N, L, v, V)$ are uniformly doubling and thus the weakly measured Gromov convergence is equivalent to the $m\text{GH}$ -convergence (see [21, Theorem 3.30 and 3.33]). \square

Corollary 1 It follows:

- (i) The stability of $\text{CD}(K, N)$ with $N \in (1, \infty]$ under \mathbf{D}_{ET} -convergence.
- (ii) The convergence of heat flows under \mathbf{D}_{ET} -convergence of $\text{CD}(K, \infty)$ spaces.
- (iii) The stability of $\text{RCD}(K, N)$ with $N \in (1, \infty]$ under \mathbf{D}_{ET} -convergence.
- (iv) The stability of the spectrum of the Laplacian under \mathbf{D}_{ET} -convergence of $\text{CD}(K, \infty)$ spaces.

Proof The proof is a direct consequence of Theorem 3 and the results contained in [21], to which we refer for the precise statements. In particular, for (i) we use [21, Theorem I and pp. 29-30], (ii) follows from [21, Theorem 5.7], for (iii) we take advantage of [21, Theorem IV], and (iv) is a consequence of [21, Theorem V]. \square

4 Limiting cases

4.1 Pure entropy distances

In the setting of Entropy-Transport problems, we call *pure entropy* problems the ones induced by the choices

$$F \in \Gamma_0(\mathbb{R}_+), \quad \mathbf{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases}$$

In this situation one can prove (see [29, Example E.5]) that for any $\mu_1, \mu_2 \in \mathcal{M}(X)$ we have

$$\text{ET}(\mu_1, \mu_2) = \inf_{\gamma \in \mathcal{M}(X)} D_F(\gamma || \mu_1) + D_F(\gamma || \mu_2) = \int_X H_0 \left(\frac{d\mu_1}{d\lambda}, \frac{d\mu_2}{d\lambda} \right) d\lambda, \quad (65)$$

where $\lambda \in \mathcal{M}(X)$ is any dominating measure of μ_1 and μ_2 and H_0 is defined as the lower semicontinuous envelope of the function

$$\tilde{H}_0(r, t) := \inf_{\theta > 0} \hat{F}(\theta, r) + \hat{F}(\theta, t). \quad (66)$$

In particular, the functional ET corresponds in this situation to the Csiszár's divergence induced by the function $s \mapsto H_0(1, s) \in \Gamma_0(\mathbb{R}_+)$ (see [14, Lemma 3]), justifying the name of pure entropy problem.

For some entropy functions F one can prove that a power a of the induced pure entropy cost ET is a distance. For instance, when $a = 1$ and $F(s) = |s - 1|$ we obtain the celebrated total variation (denoted by TV in the sequel), a distance in the space of measures inducing a strong topology. Actually, thanks to the result proved in [14, Lemma 8] and the explicit bounds contained in [31, Theorem 2.5], we know that every pure entropy distance induces the same topology of the total variation.

As shown in [14, Propositions 2, 3], we obtain another class of pure entropy distances by choosing $a = 1/2$ and the power-like entropy $F = U_p$, $p \geq 1$, defined in example 3. In this situation we have

$$H_0(r, t) = \frac{1}{p} \left[r + t - 2^{\frac{p}{p-1}} (r^{1-p} + t^{1-p})^{\frac{1}{1-p}} \right] \quad \text{if } p > 1, \quad (67)$$

$$H_0(r, t) = (\sqrt{r} - \sqrt{t})^2, \quad p = 1, \quad (68)$$

and we recognize some well-known functionals like the 2-Hellinger distance (case $p = 1$) and the triangular discrimination (case $p = 2$). We will denote these distances by PL_p .

We start with a useful lemma, valid for any pure entropy problem.

Lemma 7 Fix $a \in (0, 1]$ and let us consider the functions

$$F \in \Gamma_0(\mathbb{R}_+) \text{ such that } F'_\infty = +\infty, \quad \mathbf{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases}$$

Let us denote by PE the power a of the Entropy-Transport cost induced by F and \mathbf{c} . For any $(X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2) \in \mathbf{X}$ let us define

$$\mathbf{PE}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) := \inf \mathbf{PE}(\psi_\#^1 \mu_1, \psi_\#^2 \mu_2), \quad (69)$$

where the infimum in the right hand side is taken over all complete and separable metric spaces $(\hat{X}, \hat{\mathbf{d}})$ with isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$ and $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$.

Then

$$\mathbf{PE}^{1/a}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = \inf_C \{D_F(\gamma_1 || \mu_1) + D_F(\gamma_2 || \mu_2)\}, \quad (70)$$

where

$$C := \left\{(\gamma, \hat{\mathbf{d}}) : \gamma \in \mathcal{M}(X_1 \times X_2), \hat{\mathbf{d}} \text{ pseudo-metric coupling for } \mathbf{d}_1, \mathbf{d}_2, \text{supp}(\gamma) \subset \{\hat{\mathbf{d}} = 0\}\right\}.$$

Proof Setting

$$\ell(d) := \begin{cases} 0 & \text{if } d = 0 \\ +\infty & \text{otherwise,} \end{cases}$$

we can prove that the infimum of

$$\left\{D_F(\gamma_1 || \mu_1) + D_F(\gamma_2 || \mu_2) + \int_{X_1 \times X_2} \ell(\hat{\mathbf{d}}(x, y)) \, d\gamma\right\}^a$$

over the set

$$\tilde{C} := \{(\gamma, \hat{\mathbf{d}}) : \gamma \in \mathcal{M}(X_1 \times X_2), \hat{\mathbf{d}} \text{ pseudo-metric coupling for } \mathbf{d}_1, \mathbf{d}_2\}$$

is less or equal to the infimum in the right hand side of (69) by reasoning as in the first part of the Proof of Proposition 2. The fact that the power a of the right hand side of (70) is less or equal to the infimum as in (69) follows by noticing that

$$\int_{X_1 \times X_2} \ell(\hat{\mathbf{d}}(x, y)) \, d\gamma = \begin{cases} +\infty & \text{if } \text{supp}(\gamma) \not\subset \{\hat{\mathbf{d}} = 0\} \\ 0 & \text{otherwise.} \end{cases} \quad (71)$$

For the converse inequality, we reason in a similar way as in the proof of the point (ii) of Lemma 6. For any pseudo-metric coupling $\hat{\mathbf{d}}$ of \mathbf{d}_1 and \mathbf{d}_2 let us consider the space $((X_1 \sqcup X_2)/\sim, \hat{\mathbf{d}})$, where $x_1 \sim x_2 \iff \hat{\mathbf{d}}(x_1, x_2) = 0$. It is a complete and separable metric space as proved in Lemma 1. Denoting by $q : X_1 \sqcup X_2 \rightarrow (X_1 \sqcup X_2)/\sim$ the projection to the quotient and using the identification

$$X_1 \sqcup X_2 = X_1 \times \{0\} \cup X_2 \times \{1\},$$

we notice that

$$X_1 \times X_2 \hookrightarrow ((X_1 \sqcup X_2)/\sim) \times ((X_1 \sqcup X_2)/\sim)$$

via the injective map

$$\psi(x_1, x_2) = (\psi^1(x_1), \psi^2(x_2)) := (q(x_1, 0), q(x_2, 1)).$$

Moreover, we also have that ψ^i is an isomorphism of (X_i, d_i) into its image in $((X_1 \sqcup X_2)/\sim, \hat{d})$, $i = 1, 2$. Thus, for any measure $\gamma \in \mathcal{M}(X_1 \times X_2)$ such that $\text{supp}(\gamma) \subset \{\hat{d} = 0\}$, denoting by γ_i the marginals of γ , we can consider the measures $\psi_{\#}\gamma$ whose projections are $(\psi^1)_{\#}\gamma_1$ and $(\psi^2)_{\#}\gamma_2$. Using Lemma 4 we know that

$$D_F(\gamma_i || \mu_i) = D_F((\psi^i)_{\#}\gamma_i || (\psi^i)_{\#}\mu_i), \quad i = 1, 2$$

and the proof is completed by noticing that $\text{supp}(\psi_{\#}\gamma)$ is contained in the diagonal of the metric space $((X_1 \sqcup X_2)/\sim, \hat{d})$. \square

In the next theorem we prove that some pure entropy problems, specifically the ones generated by power-like entropies U_p , $p \geq 1$, can be recovered as a limiting case of regular Entropy-Transport problems.

Theorem 5 Fix $p \geq 1$ and let us consider the sequence of cost functions $\mathbf{c}_n = nd$ and the entropy function $F := U_p$. Let us denote by $D_{p,n}$ the Entropy-Transport distance induced by $a = 1/2$, $\mathbf{c}_n = nd$ and $F := U_p$.

Then, for every metric measure spaces (X_1, d_1, μ_1) , $(X_2, d_2, \mu_2) \in \mathbf{X}$ the limit

$$\mathbf{PL}_p((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) := \lim_{n \rightarrow \infty} D_{p,n}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) \quad \text{is well defined,} \quad (72)$$

where $D_{p,n}$ denotes the function defined as in Definition 4 upon replacing D_{ET} by $D_{p,n}$.

Moreover, $D_{p,n}$ is a regular Entropy-Transport distance and \mathbf{PL}_p defines a metric on \mathbf{X} such that

$$\mathbf{PL}_p((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) = \inf \mathbf{PL}_p(\psi_{\#}^1 \mu_1, \psi_{\#}^2 \mu_2), \quad (73)$$

where the infimum in the right hand side is taken over all complete and separable metric spaces (\hat{X}, \hat{d}) with isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$ and $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$.

Proof The first assertion follows by noticing that for any metric d we have

$$nd(x_1, x_2) \uparrow \mathbf{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2 \\ +\infty & \text{otherwise} \end{cases} \quad \text{as } n \rightarrow \infty \text{ for every } x_1, x_2 \in X. \quad (74)$$

The fact that $D_{p,n}$ is a regular Entropy-Transport distance is a consequence of [14, Theorem 6], noticing the obvious fact that nd is a complete and separable metric for any fixed n .

In particular, since for every fixed n we know that $D_{p,n}$ is a metric on \mathbf{X} by Theorem 2, we have that \mathbf{PL}_p is nonnegative, symmetric, it satisfies the triangle inequality and

$$\mathbf{PL}_p((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) = 0 \quad \text{if } (X_1, d_1, \mu_1) = (X_2, d_2, \mu_2).$$

We claim that

$$\mathbf{PL}_p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = 0$$

only if $(X_1, \mathbf{d}_1, \mu_1) = (X_2, \mathbf{d}_2, \mu_2)$ (as equivalence classes). Indeed, since $\mathbf{D}_{p,n}$ is nonnegative and nondecreasing, the fact that

$$\lim_{n \rightarrow \infty} \mathbf{D}_{p,n}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = 0$$

implies

$$\mathbf{D}_{p,n}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = 0 \quad \text{for every } n,$$

and the result follows because $\mathbf{D}_{p,n}$ is a distance on \mathbf{X} .

At this level we do not know that \mathbf{PL}_p is finite valued, which is a consequence of (73) together with the fact that \mathbf{PL}_p is a (finite valued) distance on the space of measures as recalled above.

In order to prove (73), we first notice that the monotonicity (74) easily implies that

$$\mathbf{PL}_p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \leq \inf \mathbf{PL}_p(\psi_{\sharp}^1 \mu_1, \psi_{\sharp}^2 \mu_2),$$

giving the finiteness of $\mathbf{PL}_p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2))$.

For the converse inequality, thanks to Lemma 6 we know that for every $(X_1, \mathbf{d}_1, \mu_1)$, $(X_2, \mathbf{d}_2, \mu_2)$ and for every $n \in \mathbb{N}$ there exist a measure $\gamma_n \in \mathcal{M}(X_1 \times X_2)$ and a pseudo-metric coupling $\hat{\mathbf{d}}_n$ between \mathbf{d}_1 and \mathbf{d}_2 such that:

$$\begin{aligned} \infty > \mathbf{PL}_p^2((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) &\geq \mathbf{D}_{p,n}^2((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ &= \sum_{i=1}^2 D_{U_p}(\gamma_{n,i} || \mu_i) + \int_{X_1 \times X_2} n \hat{\mathbf{d}}_n(x, y) d\gamma_n \quad \text{for every } n \in \mathbb{N}. \end{aligned} \quad (75)$$

By the superlinearity of the entropy functionals we can infer the existence (up to subsequence) of a weak limit $\gamma \in \mathcal{M}(X_1 \times X_2)$ of the sequence $\{\gamma_n\}_{n \in \mathbb{N}}$. We also know that

$$\liminf_{n \rightarrow \infty} \sum_{i=1}^2 D_{U_p}(\gamma_{n,i} || \mu_i) \geq \sum_{i=1}^2 D_{U_p}(\gamma_i || \mu_i), \quad (76)$$

where we have used the usual notation for the marginal measures. If γ is the null measure the result follows trivially. Otherwise, since the integral

$$\int_{X_1 \times X_2} \hat{\mathbf{d}}_n(x, y) d\gamma_n$$

is bounded from above we can argue as in the step 2 of the Proof of Lemma 6 and we deduce the existence of a pseudo-metric coupling $\hat{\mathbf{d}}$ between \mathbf{d}_1 and \mathbf{d}_2 such that

\hat{d}_n converges (up to subsequence) pointwise to \hat{d} and the convergence is uniform on compact sets.

By recalling the explicit formulation of the right hand side of (73) given in Lemma 7, the proof is completed if we show that $\text{supp}(\gamma) \subset \{\hat{d} = 0\}$. Let us suppose by contradiction the existence of a point $(\bar{x}, \bar{y}) \in \text{supp}(\gamma)$ such that $\hat{d}(\bar{x}, \bar{y}) = k > 0$. Fix $k/2 > r > 0$: for every $\epsilon > 0$ sufficiently small we know that there exist $m \in \mathbb{N}$ such that for every $n > m$ we have

$$\begin{aligned} \gamma_n(B_r(\bar{x}) \times B_r(\bar{y})) &\geq \gamma(B_r(\bar{x}) \times B_r(\bar{y})) - \epsilon > 0, \\ \hat{d}_n(\bar{x}, \bar{y}) - 2r &> \hat{d}(\bar{x}, \bar{y}) - 2r - \epsilon > 0. \end{aligned} \quad (77)$$

Starting from the bound in (75) we have

$$\begin{aligned} \infty &> \mathbf{PL}_p^2((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) \geq n \int_{X_1 \times X_2} \hat{d}_n(x, y) d\gamma_n \\ &\geq n \int_{B_r(\bar{x}) \times B_r(\bar{y})} \hat{d}_n(x, y) d\gamma_n \geq n \left(\hat{d}_n(\bar{x}, \bar{y}) - 2r \right) \gamma_n(B_r(\bar{x}) \times B_r(\bar{y})) \\ &\geq n \left(\hat{d}(\bar{x}, \bar{y}) - 2r - \epsilon \right) [\gamma(B_r(\bar{x}) \times B_r(\bar{y})) - \epsilon] \end{aligned}$$

that leads to a contradiction for n sufficiently large thanks to (77). \square

Definition 6 We say that a sequence of metric measure spaces $(X_n, d_n, \mu_n)_{n \in \mathbb{N}}$ strongly measured-Gromov converges to the metric measure space $(X_\infty, d_\infty, \mu_\infty)$ if there exist a complete and separable metric space (X, d) and isometric embeddings $\iota_n : X_n \rightarrow X$, $n \in \mathbb{N}$, such that $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$ in $\mathcal{M}(X)$ with respect to the total variation topology.

In the next Theorem we see that this notion of convergence coincides with the convergence induced by the distance \mathbf{PL}_p for every $p \geq 1$.

Theorem 6 Let $p \geq 1$. A sequence $(X_n, d_n, \mu_n)_{n \in \mathbb{N}}$ strongly measured-Gromov converges to $(X_\infty, d_\infty, \mu_\infty)$ if and only if

$$\mathbf{PL}_p((X_n, d_n, \mu_n), (X_\infty, d_\infty, \mu_\infty)) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (78)$$

Proof The proof is analogous to the one of Theorem 3. Let us suppose the validity of (78). By definition of \mathbf{PL}_p we know that there exist a complete and separable metric space (Y_n, d_{Y_n}) and isometric embeddings ψ_n, ψ_n^∞ of (X_n, d_n) and (X_∞, d_∞) in Y_n such that

$$\mathbf{PL}_p((\psi_n)_\# \mu_n, (\psi_\infty)_\# \mu_\infty) < \frac{1}{n}, \quad (79)$$

where PL_p is computed in the space Y_n . We now define $Y := \sqcup_{n \in \tilde{\mathbb{N}}} X_n$ endowed with the pseudo-metric d_Y

$$d_Y(y, y') := \begin{cases} d_n(y, y') & \text{if } y, y' \in X_n, n \in \tilde{\mathbb{N}} \\ d_{Y_n}(\psi_n(y), \psi_n^\infty(y')) & \text{if } y \in X_n, y' \in X_\infty \\ d_{Y_n}(\psi_n^\infty(y), \psi_n(y')) & \text{if } y \in X_\infty, y' \in X_n \\ \inf_{x \in X_\infty} d_{Y_n}(\psi_n(y), \psi_n^\infty(x)) + d_{Y_m}(\psi_m(y'), \psi_m^\infty(x)) & \text{if } y \in X_n, y' \in X_m. \end{cases}$$

We consider the space Y / \sim defined as the quotient of Y with respect to the equivalence relation

$$y \sim y' \Leftrightarrow d_Y(y, y') = 0, \quad (80)$$

and we then define the completion of this space, that we still denote by (Y, d_Y) . It is easy to see that Y is separable. By construction we notice that the set

$$\psi_n(X_n) \cup \psi_n^\infty(X_\infty) \subset Y_n$$

endowed with the distance d_{Y_n} is canonically isometrically embedded in (Y, d_Y) , so that every space $X_n, n \in \tilde{\mathbb{N}}$, is canonically isometrically embedded into Y by a map ψ'_n . We claim now that Y and ψ'_n provide a realization of the strong measured-Gromov convergence. To see this, it is enough to notice that $(\psi'_n)_\# \mu_n \rightarrow (\psi'_\infty)_\# \mu_\infty$ in $\mathcal{M}(Y)$ with respect to the topology induced by the total variation, which is a consequence of the construction of ψ'_n , (79) and the fact that PL_p induces the topology of the total variation.

For the converse, let us suppose that $(X_n, d_n, \mu_n)_{n \in \tilde{\mathbb{N}}}$ strongly measured-Gromov converges to the metric measure space $(X_\infty, d_\infty, \mu_\infty)$. By definition we know that there exist a complete and separable metric space (X, d) and isometric embeddings $\iota_n : X_n \rightarrow X, n \in \tilde{\mathbb{N}}$, such that $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$ in $\mathcal{M}(X)$ with respect to the topology of the total variation. Since PL_p metrizes this topology on $\mathcal{M}(X)$ we know that

$$\text{PL}_p((\iota_n)_\# \mu_n, (\iota_\infty)_\# \mu_\infty) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

and the result follows noticing that (X, d) is a possible competitor in the characterization of PL_p given in Theorem 5. \square

In the next easy proposition we show that the strong measured-Gromov convergence implies the weak measured-Gromov convergence.

Proposition 3 *Let $(X_n, d_n, \mu_n)_{n \in \tilde{\mathbb{N}}}$ be a sequence of metric measure spaces strong measured-Gromov converging to $(X_\infty, d_\infty, \mu_\infty)$. Then $(X_n, d_n, \mu_n)_{n \in \tilde{\mathbb{N}}}$ weakly measured-Gromov converges to $(X_\infty, d_\infty, \mu_\infty)$.*

Proof By definition there exist a complete and separable metric space (X, d) and isometric embeddings $\iota_n : X_n \rightarrow X, n \in \tilde{\mathbb{N}}$, such that $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$ in $\mathcal{M}(X)$ with respect to the total variation topology, which implies that $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$

with respect to the weak convergence. The result follows by the very definition of weak measured-Gromov convergence. \square

We conclude the section with a list of examples of convergences.

Examples 2 (1) Let us consider the metric measure space $(X_\infty, d_\infty, \mu_\infty)$ defined as the unit interval $X_\infty = [0, 1]$ endowed with the Euclidean distance and the Lebesgue measure. We know that $(X_\infty, d_\infty, \mu_\infty)$ can be approximated in the weak measured-Gromov convergence by a sequence of discrete spaces: take for instance $X_n = \{m/n\}_{m=0}^{n-1}$ endowed with the distance d_n inherited from the ambient 1-dimensional Euclidean space and the measure μ_n such that $\mu_n(m/n) = 1/n$ for every $m = 0, \dots, n-1$.

We next claim that (X_n, d_n, μ_n) does not converge to $(X_\infty, d_\infty, \mu_\infty)$ in the strong measured-Gromov convergence. Indeed, for any metric space (X, d) such that X_n is isometrically embedded in X via $\iota_n, n \in \mathbb{N}$, we have

$$\begin{aligned} \mathrm{TV}((\iota_n)_\# \mu_n, (\iota_\infty)_\# \mu_\infty) &= \sup_{A \in \mathcal{B}(X)} |(\iota_\infty)_\# \mu_\infty(A) - (\iota_n)_\# \mu_n(A)| \\ &\geq \mu_\infty \left([0, 1] \setminus \left\{ \bigcup_{m=0}^{n-1} m/n \right\} \right) = 1 \quad \text{for any } n \in \mathbb{N}. \end{aligned}$$

(2) Let us consider the metric measure space $(X_\infty, d_\infty, \mu_\infty)$ defined as the unit interval $X_\infty = [0, 1]$ endowed with the Euclidean distance and the measure $\mu_\infty = f d\mathcal{L}_{[0,1]}$, where $\mathcal{L}_{[0,1]}$ is the Lebesgue measure on $[0, 1]$. Let us define the sequence of metric measure spaces (X_n, d_n, μ_n) where $X_n = [0, 1 - 1/n]$, d_n is the Euclidean distance and $\mu_n = f_n d\mathcal{L}_{[0,1-1/n]}$. Let us suppose that $\tilde{f}_n \rightarrow f$ in $L^1([0, 1])$, where

$$\tilde{f}_n(x) = \begin{cases} f_n(x) & \text{if } 0 \leq x \leq 1 - 1/n \\ 0 & \text{if } 1 - 1/n < x \leq 1. \end{cases}$$

Then, $(X_n, d_n, \mu_n) \rightarrow (X_\infty, d_\infty, \mu_\infty)$ in the strong measured-Gromov convergence. To see this, it is enough to notice that for every $n \in \mathbb{N}$ the maps $\iota_n : X_n \rightarrow X_\infty$ defined as $\iota_n(x) = x$ provides an isometric embedding such that the convergence $(\iota_n)_\# \mu_n \rightarrow (\iota_\infty)_\# \mu_\infty$ with respect the total variation distance is exactly equivalent to $\tilde{f}_n \rightarrow f$ in $L^1([0, 1])$.

(3) Let (X_n, d_n, μ_n) be the sequence of collapsing flat tori $S^1 \times \frac{1}{n} S^1 \subset \mathbb{R}^4$ endowed with the normalized measures $\mu_n := n/(4\pi^2) \mathrm{dvol}_{S^1 \times \frac{1}{n} S^1}$. It is a standard fact that (X_n, d_n, μ_n) converges to $(X_\infty, d_\infty, \mu_\infty) = (S^1, d_{S^1}, (2\pi)^{-1} \mathcal{L}^1)$ in the weak measured-Gromov sense (this a standard example of a *collapsing sequence*).

We claim that the convergence cannot be improved to strong measured-Gromov. Indeed, for any metric space (X, d) such that X_n is isometrically embedded in X

via $\iota_n, n \in \tilde{\mathbb{N}}$, we have

$$\begin{aligned} \mathrm{TV}((\iota_n)_\# \mu_n, (\iota_\infty)_\# \mu_\infty) &= \sup_{A \in \mathcal{B}(X)} |(\iota_\infty)_\# \mu_\infty(A) - (\iota_n)_\# \mu_n(A)| \\ &\geq \mu_n \left(\left(S^1 \times \frac{1}{n} S^1 \right) \setminus \gamma_n(S^1) \right) = 1 \quad \text{for any } n \in \mathbb{N}, \end{aligned}$$

where $\gamma_n : S^1 \rightarrow S^1 \times \frac{1}{n} S^1$ is an arbitrary isometric immersion.

4.2 Sturm's distances

We notice that the classical p -Wasserstein distance $\mathcal{W}_p, p \geq 1$, can be recovered as a particular case of Entropy-Transport problem with the choices

$$F(s) = I_1(s) := \begin{cases} 0 & \text{if } s = 1 \\ +\infty & \text{otherwise} \end{cases} \quad \mathbf{c}(x_1, x_2) = d^p(x_1, x_2). \quad (81)$$

It is clear that \mathcal{W}_p is not a *regular* Entropy-Transport distance, however we show now that we can recover the \mathbf{D}_p -distance of Sturm (defined in Definition 3) as a limiting case of our framework.

Theorem 7 Fix $p \geq 1$ and let us consider the cost function $\ell(d) := d^p$, the entropy function $F := U_1$, and the power $a := 1/p$. Let us denote by $\mathbf{D}_{\mathrm{ET},n}$ the power a of the Entropy-Transport cost induced by $F_n := nU_1$ and $\mathbf{c} = \ell(d)$. Then, for every metric measure spaces $(X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2) \in \mathbf{X}$

$$\mathcal{D}_p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) := \lim_{n \rightarrow \infty} \mathbf{D}_{\mathrm{ET},n}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \quad \text{is well defined,} \quad (82)$$

where $\mathbf{D}_{\mathrm{ET},n}$ denotes the function defined as in Definition 4 upon replacing \mathbf{D}_{ET} by $\mathbf{D}_{\mathrm{ET},n}$. Moreover, for every metric measure spaces $(X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2) \in \mathbf{X}_{1,p}$ we have

$$\mathcal{D}_p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = \mathbf{D}_p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \quad p \geq 1. \quad (83)$$

Proof We start by proving that the limit (82) exists on the set \mathbf{X} . To see this, we notice that $nF(s) \uparrow I_1(s)$ for every $s \in [0, \infty)$. In particular, using the explicit formulation of $\mathbf{D}_{\mathrm{ET},n}$ proved in Proposition 2 (we remark that we have not used the fact that \mathbf{D} is a distance in the proof of the proposition), we can infer that $\mathbf{D}_{\mathrm{ET},n}$ is nondecreasing so that the limit exists.

It remains to prove that for every $p \geq 1$ we have $\mathcal{D}_p = \mathbf{D}_p$ on the set $\mathbf{X}_{1,p}$. Since for every $n \in \mathbb{N}$, for every complete and separable metric space (X, \mathbf{d}) and for every $\mu, \gamma \in \mathcal{M}(X)$ we have

$$D_{F_n}(\gamma || \mu) \leq D_{I_1}(\gamma || \mu),$$

it is clear that $\mathcal{D}_p \leq \mathbf{D}_p$. For the converse inequality, we know that for every $(X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2) \in \mathbf{X}_{1,p}$ we have

$$\begin{aligned} \mathbf{D}_p^p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) &\geq \mathcal{D}_p^p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ &\geq \mathbf{D}_{\text{ET},n}^p((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ &= \sum_{i=1}^2 D_{F_n}(\gamma_{n,i} || \mu_i) + \int_{X_1 \times X_2} \hat{\mathbf{d}}_n^p(x, y) d\gamma_n \end{aligned} \quad (84)$$

for some $\gamma_n \in \mathcal{M}(X_1 \times X_2)$ and metric coupling $\hat{\mathbf{d}}_n$ of \mathbf{d}_1 and \mathbf{d}_2 , whose existence is a consequence of Lemma 6 (notice that we have only used the properties of the cost and the entropy in the proof of the lemma, while the fact that \mathbf{D}_{ET} is a metric plays no role). Since D_{F_n} is bounded from above by the superlinear entropy D_{I_1} , by using Lemma 2 and Lemma 3 we can infer that γ_n is weakly converging (up to subsequence) to a limit $\gamma \in \mathcal{M}(X_1 \times X_2)$ and

$$\liminf \sum_{i=1}^2 D_{F_n}(\gamma_{n,i} || \mu_i) \geq \sum_{i=1}^2 D_{I_1}(\gamma_i || \mu_i).$$

By reasoning as in step 2 and step 3 of Lemma 6, we know that there exists a pseudo-metric coupling $\hat{\mathbf{d}}$ of \mathbf{d}_1 and \mathbf{d}_2 such that $\hat{\mathbf{d}}_n$ converges (up to subsequence) pointwise to $\hat{\mathbf{d}}$ and the convergence is uniform on compact sets. Moreover, we have that

$$\liminf_n \int_{X_1 \times X_2} \hat{\mathbf{d}}_n^p(x, y) d\gamma_n \geq \int_{X_1 \times X_2} \hat{\mathbf{d}}^p(x, y) d\gamma, \quad (85)$$

and the result follows since $\hat{\mathbf{d}}$ and γ are competitors in the explicit formulation of \mathbf{D}_p (see [43, Lemma 3.3]) as a consequence of the fact that

$$D_{I_1}(\gamma_i || \mu_i) < \infty \iff \gamma_i = \mu_i, \quad i = 1, 2.$$

□

Remark 4 We point out that we are not claiming that the sequence $\mathbf{D}_{\text{ET},n}$ defined in Theorem 7 is a sequence of *regular* Entropy-Transport distances. Actually, this is the case for $p = 2$ as a consequence of [29, Theorem 7.25], noticing that the cost of the Entropy-Transport problem induced by (nU_1, \mathbf{d}^2) is n times the cost of the Entropy-Transport problem induced by $(U_1, (\mathbf{d}/\sqrt{n})^2)$ and \mathbf{d}/\sqrt{n} is trivially a complete and separable distance.

In this situation, one can show that \mathbf{D}_2 defines a metric possibly attaining the value $+\infty$ on the whole set \mathbf{X} , by reasoning as in the Proof of Theorem 5.

4.3 Piccoli-Rossi distance

A natural extension of the \mathcal{W}_1 -metric in the context of Entropy-Transport problem is the Piccoli-Rossi *generalized Wasserstein distance* BL [35, 36], induced by the choices

$$F(s) = |s - 1|, \quad \mathbf{c}(x_1, x_2) = \mathbf{d}(x_1, x_2). \quad (86)$$

We notice that the entropy function is not superlinear.

It is proved in [35] that BL is a complete distance on $\mathcal{M}(X)$ for every Polish space (X, \mathbf{d}) ([35] is in the Euclidean setting, however the proof for a Polish space can be performed verbatim).

By exploiting the dual formulation of this distance, we know that BL corresponds to the so-called *flat metric* or *bounded Lipschitz distance* (see [36, Theorem 2]), namely

$$\begin{aligned} \text{BL}(\mu_1, \mu_2) &= \sup \left\{ \int_X f \, \mathbf{d}(\mu_1 - \mu_2) : \|f\|_\infty \leq 1, \|f\|_{\text{Lip}} \leq 1 \right\} \quad \text{for any } \mu_1, \mu_2 \in \mathcal{M}(X). \end{aligned} \quad (87)$$

We also recall this useful lemma, which is proved in [35, Proposition 1].

Lemma 8 *Given $\mu_1, \mu_2 \in \mathcal{M}(X)$, let us consider the Entropy-Transport problem induced by (F, \mathbf{c}) defined in (86). Then the infimum of the problem (26) is attained by a measure $\gamma \in \mathcal{M}(X \times X)$ such that $\gamma_i := (\pi^i)_\# \gamma \leq \mu_i$, $i = 1, 2$.*

We have the following:

Theorem 8 *Fix $a = 1$, $\ell(d) := d$ and let us consider the sequence $(F_n)_{n \geq 2}$ defined by*

$$F_n(s) := \begin{cases} |s - 1| & \text{if } 0 \leq s \leq n \\ \frac{(s-1)^2}{n-1} & \text{if } s > n. \end{cases}$$

Let us denote by $D_{\text{ET},n}$ the Entropy-Transport cost induced by a , F_n and $\mathbf{c} = \ell(d)$.

Then, for every metric measure spaces $(X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2) \in \mathbf{X}$ the quantity

$$\begin{aligned} \text{BL}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ := \mathbf{D}_{\text{ET},n}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \quad \text{is well defined,} \end{aligned} \quad (88)$$

where $\mathbf{D}_{\text{ET},n}$ denotes the function defined as in Definition 4 upon replacing D_{ET} by $D_{\text{ET},n}$.

Moreover, BL defines a complete metric on \mathbf{X} such that

$$\text{BL}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) = \inf \text{BL}(\psi_\#^1 \mu_1, \psi_\#^2 \mu_2), \quad (89)$$

where the infimum in the right hand side is taken over all complete and separable metric spaces (\hat{X}, \hat{d}) with isometric embeddings $\psi^1 : \text{supp}(\mu_1) \rightarrow \hat{X}$ and $\psi^2 : \text{supp}(\mu_2) \rightarrow \hat{X}$.

Proof We notice that $(F_n)_{n \geq 2}$ is a sequence of continuous superlinear entropy functions. We also know that $F_n(s) = |s - 1|$ in $[0, 1]$ and $F_n(s) \geq |s - 1|$ in $[0, \infty)$ for every $n \geq 2$, which implies that $D_{\text{ET},n}$ coincide with BL thanks to Lemma 8. In particular we see that $D_{\text{ET},n}$ does not depend on n and also the identity (89) follows.

The fact that BL is a complete distance on \mathbf{X} is a consequence of the completeness of BL (and thus $D_{\text{ET},n}$) on the set of measures $\mathcal{M}(X)$, and can be proved along the lines of Step 2 in the Proof Theorem 2. \square

Remark 5 We observe that the sequence $D_{\text{ET},n}$ defined in Theorem 8 is not a sequence of regular Entropy-Transport distances. The problem here is that the topology induced by the distance BL does not coincide with the weak topology, but it requires an additional tightness condition (see [35, Theorem 3] for all the details).

4.4 Bounds between distances

The aim of this last short section is to give some explicit bounds between the distances discussed in the paper.

Proposition 4 Let us denote by HK, GHK, QPL_p (for $1 < p \leq 3$) and LPL_p (for $p > 1$) the regular Entropy-Transport distances defined in examples (1), (2), (3) and (4) respectively. Accordingly, we denote by D_{HK} , D_{GHK} , D_{QPL_p} and D_{LPL_p} the induced Sturm-Entropy-Transport distances. The following inequalities hold:

- (1) $D_{\text{GHK}} \leq D_{\text{HK}}$.
- (2) $D_{QPL_p} \leq D_{\text{GHK}} \leq \sqrt{p} D_{QPL_p} \quad 1 < p \leq 3$.
- (3) $D_{LPL_p} \leq PL_p \quad p > 1$.

Moreover, for every regular entropy transport distance D_{ET} induced by $(1/p, F, \ell)$ where $p \geq 1$, $F \in \Gamma_0(\mathbb{R}_+)$, $\ell(d) = d^p$ we have:

- (4) $D_{\text{ET}} \leq D_p \quad p \geq 1$.

Proof (1) is a consequence of the bound proved in [29, Sect. 7.8].

(2) follows by the corresponding inequality proved in [14, Proposition 7].

(3) has been shown along the lines of the Proof of Theorem 5 (notice that D_{LPL_p} equals $D_{p,1}$ in the notation of that Theorem).

(4) is a consequence of the explicit formulations of D_p and D_{ET} , by noticing that for any $F \in \Gamma_0(\mathbb{R}_+)$ we have $F \leq I_1$ where I_1 has been defined in (81). \square

5 Comparison with conic Gromov–Wasserstein

Let D_{ET} be a regular entropy transport distance induced by (a, F, ℓ) . Recalling the construction introduced in Remark 1, given F and a number $c \geq 0$ we can associate

to the Entropy-Transport problem a function $H_c : [0, +\infty) \times [0, +\infty) \rightarrow [0, +\infty]$ called marginal perspective function. Moreover, for any complete and separable metric space (X, d) the function

$$H : \mathfrak{C}(X) \times \mathfrak{C}(X) \rightarrow [0, +\infty], \quad H([x, r]; [y, s]) := H_{\ell(d(x,y))}(r, s),$$

is such that H^a is a distance on $\mathfrak{C}(X)$. In particular, we have

$$H_{\ell(w_3)}^a(r, t) \leq H_{\ell(w_1)}^a(r, s) + H_{\ell(w_2)}^a(s, t), \quad \text{for any } r, s, t \in [0, \infty), \quad (90)$$

and for any $w_1, w_2, w_3 \in [0, \infty)$ such that there exists a complete and separable metric space (X, d) with $w_1 = d(x_1, x_2)$, $w_2 = d(x_2, x_3)$, $w_3 = d(x_1, x_3)$, $x_1, x_2, x_3 \in X$.

We also recall that H is positively 1-homogeneous in the scalar variables, i.e.

$$H_c(\lambda r, \lambda s) = \lambda H_c(r, s) \quad \text{for any } \lambda \geq 0, c \geq 0, r, s \in [0, \infty).$$

Let X be a metric space and fix $\bar{x} \in X$. We define the canonical projection $p : X \times [0, \infty) \rightarrow \mathfrak{C}(X)$ as $p(x, r) = [x, r]$. We also introduce the maps

$$r : \mathfrak{C}(X) \rightarrow [0, +\infty), \quad r[x, r] := r, \quad (91)$$

$$x : \mathfrak{C}(X) \rightarrow X, \quad x[x, r] := \begin{cases} x & \text{if } r > 0, \\ \bar{x} & \text{if } r = 0. \end{cases} \quad (92)$$

We denote by $\mathfrak{h} = (\mathfrak{h}_1, \mathfrak{h}_2) = ([x_1, r_1], [x_2, r_2])$ a point on $\mathfrak{C}(X_1) \times \mathfrak{C}(X_2)$, and we set $r_i(\mathfrak{h}) := r(\mathfrak{h}_i)$, $x_i(\mathfrak{h}) := x(\mathfrak{h}_i)$.

Given $p \geq 1$, the p -homogeneous marginals of a measure $\alpha \in \mathcal{M}(\mathfrak{C}(X_1) \times \mathfrak{C}(X_2))$ are defined as

$$\mathfrak{h}_i^p(\alpha) := (x_i)_\#(r_i^p \alpha), \quad i = 1, 2.$$

Following the approach of [39], we define the conic Gromov–Wasserstein distance CGW^a between two metric measure spaces (X_1, d_1, μ_1) , (X_2, d_2, μ_2) , as the power a of

$$\text{CGW}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)) := \inf_{\alpha \in \mathcal{U}_p(\mu_1, \mu_2)} \mathcal{H}(\alpha) \quad (93)$$

where

$$\mathcal{H}(\alpha) := \int \int H_{\ell(|d_1(x, x') - d_2(y, y')|)}((rr')^p, (ss')^p) d\alpha([x, r], [y, s]) d\alpha([x', r'], [y', s']), \quad (94)$$

and

$$\mathcal{U}_p(\mu_1, \mu_2) := \{\alpha \in \mathcal{M}(\mathfrak{C}(X_1) \times \mathfrak{C}(X_2)) : \mathfrak{h}_i^p(\alpha) = \mu_i, i = 1, 2\}. \quad (95)$$

In [39], the following main result has been obtained:

Theorem 9 ([39, Theorem 1]) *If H^a is a distance on the cone then the conic Gromov–Wasserstein distance CGW^a is a metric on \mathbf{X} .*

As we will see, it is possible to prove an inequality between the conic Gromov–Wasserstein distance and the Sturm–Entropy–Transport distance. We start with a lemma.

Lemma 9 *Let D_{ET} be a regular entropy transport distance induced by (a, F, ℓ) and let $p \geq 1$. Let $(X_1, d_1), (X_2, d_2)$ be two complete and separable metric spaces and let \hat{d} be a pseudo-metric coupling between d_1 and d_2 . For any $\mu_1 \in \mathcal{M}(X_1)$ and $\mu_2 \in \mathcal{M}(X_2)$ we have*

$$\begin{aligned} & \min_{\alpha \in \mathcal{U}_p(\mu_1, \mu_2)} \int_{\mathfrak{C}(X_1) \times \mathfrak{C}(X_2)} H_{\ell(\hat{d}(x, y))}(r^p, s^p) d\alpha([x, r], [y, s]) \\ &= \min_{\gamma \in \mathcal{M}(X_1 \times X_2)} \sum_{i=1}^2 D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{d}(x, y)) d\gamma. \end{aligned} \quad (96)$$

In particular, for any pair of metric measure spaces $(X_1, d_1, \mu_1), (X_2, d_2, \mu_2)$ there exists $\alpha \in \mathcal{U}_p(\mu_1, \mu_2)$ such that

$$\begin{aligned} & \int_{\mathfrak{C}(X_1) \times \mathfrak{C}(X_2)} H_{\ell(\hat{d}(x, y))}(r^p, s^p) d\alpha([x, r], [y, s]) \\ &= \mathbf{D}_{\text{ET}}^{1/a}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2)), \end{aligned} \quad (97)$$

where \hat{d} is an optimal pseudo-metric coupling for $\mathbf{D}_{\text{ET}}((X_1, d_1, \mu_1), (X_2, d_2, \mu_2))$.

Proof Notice that under our assumptions we are in the basic *coercive* setting described in [29, Sect. 3.1]. The equality stated in (96), corresponding to the equivalence between the homogeneous formulation based on the function H and the primal Entropy–Transport formulation, is thus a consequence of [29, Theorem 5.8 (iii)] and the use of the projection map

$$\begin{aligned} \mathfrak{p} : X_1 \times \mathbb{R}_+ \times X_2 \times \mathbb{R}_+ &\rightarrow \mathfrak{C}(X_1) \times \mathfrak{C}(X_2), \\ \mathfrak{p}(x, r, y, s) &:= (\mathfrak{p} \otimes \mathfrak{p})((x, r), (y, s)) = ([x, r], [y, s]) \end{aligned}$$

for passing to the cone (see [29, Sect. 7] for all the details in the case of the Hellinger–Kantorovich distance, the general case follows straightforwardly).

Once (96) has been proved, (97) can be deduced by recalling the characterization of \mathbf{D}_{ET} given in the point (i) of Lemma 6. \square

We can now state the main result of this section. It should be compared with [32, Theorem 5.1], [45, Proposition 2.6].

Proposition 5 *Let D_{ET} be a regular entropy transport distance induced by (a, F, ℓ) and let CGW^a be the associated conic Gromov–Wasserstein distance. For any pair of metric measure spaces $(X_1, \mathbf{d}_1, \mu_1)$, $(X_2, \mathbf{d}_2, \mu_2)$, it holds*

$$\begin{aligned} & \text{CGW}^a((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ & \leq (\mu_1(X_1)^a + \mu_2(X_2)^a) \mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)). \end{aligned} \quad (98)$$

Proof By taking advantage of Lemma 6, let us consider an optimal pseudo-metric coupling $\hat{\mathbf{d}}$ for $\mathbf{D}_{\text{ET}}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2))$. Thanks to Lemma 9, let us also consider $\alpha \in \mathcal{U}_p(\mu_1, \mu_2)$ satisfying

$$\int H_{\ell(\hat{\mathbf{d}}(x,y))}(r^p, s^p) \, \mathrm{d}\alpha([x, r], [y, s]) = \mathbf{D}_{\text{ET}}^{1/a}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)). \quad (99)$$

To shorten the notation, let us set $\boldsymbol{\eta} := ([x, r], [y, s])$ and $\boldsymbol{\eta}' := ([x', r'], [y', s'])$. By the triangle inequality for $\hat{\mathbf{d}}$, and the fact that $\ell(\cdot)$ and $H(r, t)$ are increasing for any r, t , we have

$$\begin{aligned} & \text{CGW}^a((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)) \\ & \leq \left(\int \int H_{\ell(\hat{\mathbf{d}}(x,y)+\hat{\mathbf{d}}(x',y'))}((rr')^p, (ss')^p) \, \mathrm{d}\alpha(\boldsymbol{\eta}) \, \mathrm{d}\alpha(\boldsymbol{\eta}') \right)^a \\ & \leq \left(\int \int \left[H_{\ell(\hat{\mathbf{d}}(x,y))}^a((rr')^p, (r's')^p) + H_{\ell(\hat{\mathbf{d}}(x',y'))}^a((r's')^p, (ss')^p) \right]^{\frac{1}{a}} \, \mathrm{d}\alpha(\boldsymbol{\eta}) \, \mathrm{d}\alpha(\boldsymbol{\eta}') \right)^a \\ & \leq \left(\int \int H_{\ell(\hat{\mathbf{d}}(x,y))}(r^p, s^p)(r')^p \, \mathrm{d}\alpha(\boldsymbol{\eta}) \, \mathrm{d}\alpha(\boldsymbol{\eta}') \right)^a \\ & \quad + \left(\int \int H_{\ell(\hat{\mathbf{d}}(x',y'))}((r')^p, (s')^p) s^p \, \mathrm{d}\alpha(\boldsymbol{\eta}) \, \mathrm{d}\alpha(\boldsymbol{\eta}') \right)^a \end{aligned} \quad (100)$$

where we have also used the Minkowski inequality and the homogeneity of H in the last passage, and we have taken advantage of (90) with $w_3 = \hat{\mathbf{d}}(x, y) + \hat{\mathbf{d}}(x', y')$, $w_1 = \hat{\mathbf{d}}(x, y)$ and $w_2 = \hat{\mathbf{d}}(x', y')$. To justify the use of (90) we can argue as follows: if $\hat{\mathbf{d}}(x, y) = 0$ or $\hat{\mathbf{d}}(x', y') = 0$ there is nothing to prove, otherwise we notice that the three points metric space $(\{A, B, C\}, \tilde{\mathbf{d}})$ with mutual distances between different points defined as $\tilde{\mathbf{d}}(A, B) := \hat{\mathbf{d}}(x, y)$, $\tilde{\mathbf{d}}(B, C) := \hat{\mathbf{d}}(x', y')$, $\tilde{\mathbf{d}}(A, C) := \hat{\mathbf{d}}(x, y) + \hat{\mathbf{d}}(x', y')$ is indeed a complete and separable metric space for any $x, x' \in X_1, y, y' \in X_2$.

Using the definition of $\alpha \in \mathcal{U}_p(\mu_1, \mu_2)$, we can now perform the integrals in (100) obtaining

$$\begin{aligned} & \int \int H_{\ell(\hat{\mathbf{d}}(x,y))}(r^p, s^p)(r')^p \, \mathrm{d}\alpha([x, r], [y, s]) \, \mathrm{d}\alpha([x', r'], [y', s']) \\ & = \mu_1(X) \int H_{\ell(\hat{\mathbf{d}}(x,y))}(r^p, s^p) \, \mathrm{d}\alpha([x, r], [y, s]) \\ & = \mu_1(X_1) \mathbf{D}_{\text{ET}}^{1/a}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)), \end{aligned}$$

and similarly

$$\begin{aligned} & \int \int H_{\ell(\hat{d}(x', y'))} ((r')^p, (s')^p) s^p \, d\alpha([x, r], [y, s]) \, d\alpha([x', r'], [y', s']) \\ &= \mu_2(X_2) \mathbf{D}_{\text{ET}}^{1/a}((X_1, \mathbf{d}_1, \mu_1), (X_2, \mathbf{d}_2, \mu_2)). \end{aligned}$$

We thus reach the desired conclusion. \square

Acknowledgements The project started when N.D.P. was visiting A.M. in the fall 2018 at the Mathematics Institute of the University of Warwick, and took advantage of a second visit of N.D.P. to the Mathematical Institute of the University of Oxford in March 2020. The authors wish to thank both the institutions for the inspiring atmosphere and the excellent working conditions. A.M. acknowledges the support of the EPSRC First Grant EP/R004730/1 “Optimal transport and geometric analysis” and of the European Research Council (ERC), under the European’s Union Horizon 2020 research and innovation programme, via the ERC Starting Grant “CURVATURE”, grant agreement No. 802689. The authors wish to warmly thank Giuseppe Savaré for valuable discussions on the topics of the paper and are grateful to the anonymous reviewers, for their suggestions and comments that helped to improve a previous version of the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alvarez-Melis, D., Jaakkola, T.S.: “Gromov-Wasserstein Alignment of Word Embedding Spaces”, Proc. 2018 Conference on Empirical Methods in Natural Language Processing, pp. 1881–1890 (2018)
2. Ambrosio, L.: “Calculus, heat flow and curvature-dimension bounds in metric measure spaces”, In: Proceedings of the ICM 2018, Rio de Janeiro, Vol. 1, pp. 301–340
3. Ambrosio, L., Gigli, N., Mondino, A., Rajala, T.: Riemannian Ricci curvature lower bounds in metric measure spaces with σ -finite measure. Trans. Amer. Math. Soc. **367**(7), 4661–4701 (2015)
4. Ambrosio, L., Gigli, N., Savaré, G.: Metric measure spaces with Riemannian Ricci curvature bounded from below. Duke Math. J. **163**, 1405–1490 (2014)
5. Ambrosio, L., Mondino, A., Savaré, G.: “Nonlinear diffusion equations and curvature conditions in metric measure spaces”, Mem. Amer. Math. Soc., 262(1270) (2019)
6. Bunne, C., Alvarez-Melis, D., Krause, A., Jegelka, S.: “Learning generative models across incompatible spaces”, Proc. 36th Intern. Conference on Machine Learning, Long Beach, California, PMLR 97 (2019)
7. Cavalletti, F., Milman, E.: The Globalization Theorem for the Curvature Dimension Condition. Invent. Math. **226**, 1–137 (2021)
8. Cheeger, J., Colding, T.: On the structure of spaces with Ricci curvature bounded below I. J. Diff. Geom. **45**, 406–480 (1997)
9. Chizat, L., Bach, F.: On the global convergence of gradient descent for over-parameterized models using optimal transport. Adv. Neural. Inf. Process. Syst. **31**, 3036–3046 (2018)
10. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Unbalanced optimal transport: dynamic and Kantorovich formulations. J. Funct. Anal. **274**(11), 3090–3123 (2018)
11. Chowdhury, S., Mémoi, F.: The Gromov-Wasserstein distance between networks and stable network invariants. Information and Inference: A Journal of the IMA **8**(4), 757–787 (2019)

12. Csiszár, I.: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8**, 85–108 (1963)
13. Demetci, P., Santorella, R., Sandstede, B., Noble, W. S., Singh, R.: “Gromov–Wasserstein Optimal Transport to Align Single-Cell Multi-Omics Data”, 2020 ICML Workshop on Computational Biology, paper 51 (2020)
14. De Ponti, N.: Metric properties of homogeneous and spatially inhomogeneous F -divergences. *IEEE Trans. Inf. Theory* **66**(5), 2872–2890 (2020)
15. De Ponti, N.: “Optimal transport: entropic regularizations, geometry and diffusion PDEs”, Phd Thesis, <http://cvgmt.sns.it/paper/4525/> (2019)
16. Erbar, M., Kuwada, K., Sturm, K.T.: On the Equivalence of the Entropic Curvature-Dimension Condition and Bochner’s Inequality on Metric Measure Space. *Invent. Math.* **201**(3), 993–1071 (2015)
17. Feydy, J., Roussillon, P., Trounev, A., Gori, P.: “Fast and scalable optimal transport for brain trac-tograms”, Intern. Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 636–644. Springer (2019)
18. Frogner, C., Zhang, C., Mobahi, H., Araya, M., Poggio, T.A.: Learning with a Wasserstein loss. *Adv. Neural. Inf. Process. Syst.* **28**, 2053–2061 (2015)
19. Fukaya, K.: Collapsing of Riemannian manifolds and eigenvalues of Laplace operator. *Invent. Math.* **87**(3), 517–547 (1987)
20. Gigli, N.: On the differential structure of metric measure spaces and applications. *Mem. Amer. Math. Soc.* **236**(1113), vi+91 (2015)
21. Gigli, N., Mondino, A., Savaré, G.: Convergence of pointed non-compact metric measure spaces and stability of Ricci curvature bounds and heat flows. *Proc. London Math. Soc.* (3) **111**, 1071–1129 (2015)
22. Gigli, N., Pasqualetto, E.: Lectures on Nonsmooth Differential Geometry. In: SISSA Springer Series, vol. 2. Springer International Publishing, New York (2020)
23. Gromov, M.: Metric Structures for Riemannian and Non-riemannian Spaces. In: Progress in Mathematics, vol. 152. Birkhäuser Boston Inc, Boston (1999) In: vol. 152. Birkhäuser Boston Inc, Boston (1999) Progress in Mathematics, vol. 152. Birkhäuser Boston Inc, Boston (1999)
24. Kawano, S., Mason, J.K.: Classification of atomic environments via the Gromov–Wasserstein distance. *Comp. Mater. Sci.* **188**, 110144 (2021)
25. Kondratyev, S., Monsaingeon, L., Vorotnikov, D.: A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations* **21**(11/12), 1117–1164 (2016)
26. Ledoux, M.: The Concentration of Measure Phenomenon. In: Math. Surveys and Monographs, vol. 89. American Math. Soc, United States (2001)
27. Lee, J., Bertrand, N.P., Rozell, C.J.: Unbalanced Optimal Transport Regularization for Imaging Problems. *IEEE Trans. Comput. Imaging* **6**, 1219–1232 (2020)
28. Liero, M., Mielke, A., Savaré, G.: Optimal Transport in Competition with Reaction: The Hellinger–Kantorovich Distance and Geodesic Curves. *SIAM J. Math. Analysis* **48**(4), 2869–2911 (2016)
29. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Invent. Math.* **211**(3), 969–1117 (2018)
30. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.* **169**, 903–991 (2009)
31. Luise, G., Savaré, G.: “Contraction and regularizing properties of heat flows in metric measure spaces”, Discrete and Continuous Dynamical Systems Series S, early access, <https://doi.org/10.3934/dcdss.2020327> (2020)
32. Mémoi, F.: Gromov–Wasserstein distances and the metric approach to object matching. *Found. Comput. Math.* **11**(4), 417–487 (2011)
33. Peyré, G., Cuturi, M.: Computational Optimal Transport: With Applications to Data Science. *Found. Trend Mach. Learning* **11**(5–6), 355–607 (2019)
34. Peyré, G., Cuturi, M., Solomon, J.: “Gromov–Wasserstein averaging of kernel and distance matrices”, ICML’16: Proceedings of the 33rd International Conference on International Conference on Machine Learning, Vol. 48, pp. 2664–2672 (2016)
35. Piccoli, B., Rossi, F.: Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.* **211**, 335–358 (2014)
36. Piccoli, B., Rossi, F.: On properties of the Generalized Wasserstein distance. *Arch. Ration. Mech. Anal.* **222**, 1339–1365 (2016)

37. Rotskoff, G., Jelassi, S., Bruna, J., Vanden-Eijnden, E.: "Global convergence of neuron birth-death dynamics", Proc. 36th Intern. Conference on Machine Learning, Long Beach, California, PMLR 97 (2019)
38. Séjourné, T., Feydy, J., Vialard, F.-X., Trounev, A., Peyré, G.: "Sinkhorn Divergences for Unbalanced Optimal Transport", preprint [arXiv:1910.12958](https://arxiv.org/abs/1910.12958)
39. Séjourné, T., Vialard, F.X., Peyré, G.: "The Unbalanced Gromov Wasserstein Distance: Conic Formulation and Relaxation", Advances in Neural Information Processing Systems 34, NeurIPS 2021 (2021)
40. Schmitzer, B., Schnörr, C.: Modelling convex shape priors and matching based on the Gromov–Wasserstein distance. *J. Math. Imaging and Vision* **46**(1), 143–159 (2013)
41. Shioya, T.: "Metric measure geometry", Vol. 25, IRMA Lectures in Math. and Theoretical Phys., EMS Publishing House, Zürich, pp. xi+182 (2016)
42. Solomon, J., Peyré, G., Kim, V.G., Sra, S.: Entropic metric alignment for correspondence problems. *ACM Trans. Gr.* **35**(4), 1–13 (2016)
43. Sturm, K.T.: On the geometry of metric measure spaces. I. *Acta Math.* **196**(1), 65–131 (2006)
44. Sturm, K.T.: On the geometry of metric measure spaces. II. *Acta Math.* **196**(1), 133–177 (2006)
45. Sturm, K.T.: "The space of spaces: curvature bounds and gradient flows on the space of metric measure spaces", preprint [arXiv:1208.0434](https://arxiv.org/abs/1208.0434)
46. Xu, H., Luo, D., Carin, L.: Scalable Gromov–Wasserstein Learning for Graph Partitioning and Matching. *Adv. Neural. Inf. Process. Syst.* **32**, 3046–3056 (2019)
47. Vayer, T., Chapel, L., Flamary, R., Tavenard, R., Courty, N.: Fused Gromov–Wasserstein Distance for Structured Objects. *Algorithms* **13**(9), 212 (2020)
48. Villani, C.: *Optimal Transport - Old and New*. Springer, Berlin (2009)
49. Weitkamp, C. A., Proksch, K., Tameling, C., Munk, A.: "Gromov-Wasserstein Distance based Object Matching: Asymptotic Inference" preprint [arXiv:2006.12287](https://arxiv.org/abs/2006.12287) (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.